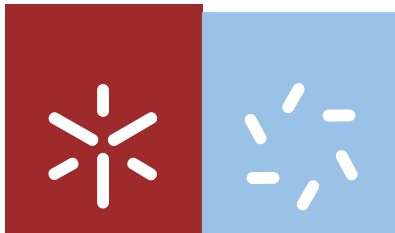


Universidade do Minho
Escola de Ciências

Davide Ribeiro da Cruz

Desenvolvimento e validação de metodologias computacionais de desenho de fármacos baseado em fragmentos químicos



Universidade do Minho

Escola de Ciências

Davide Ribeiro da Cruz

**Desenvolvimento e validação de
metodologias computacionais de
desenho de fármacos baseado em
fragmentos químicos**

Dissertação de Mestrado
Mestrado em Química Medicinal

Trabalho realizado sob a orientação do
Professor Doutor Nuno Miguel da Silva Micaêlo

Outubro de 2012

DECLARAÇÃO

Nome: Davide Ribeiro da Cruz

Endereço eletrónico: pg16398@alunos.uminho.pt

Título da tese de mestrado: Desenvolvimento e validação de metodologias computacionais de desenho de fármacos baseado em fragmentos químicos

Orientador(es): Nuno Miguel da Silva Micaêlo

Ano de conclusão: 2012

Designação do Mestrado: Mestrado em Química Medicinal

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ____/____/____

Assinatura: _____

Agradecimentos

Agradeço ao meu orientador, por me ter acolhido e acompanhado desde o primeiro dia no grupo de investigação e no qual descobri o que é fazer ciência, e que a modelação molecular é mais do que visualizar moléculas e fazer imagens bonitas. Agradeço também por apostar no meu futuro e na minha formação.

Agradeço aos membros do grupo, à Tarsila, António, João e Bárbara, pelo apoio que me deram, e pelo ambiente de trabalho que proporcionaram. Agradeço especialmente à Tarsila, pelo apoio dado, e pela ajuda na prestada no decorrer deste trabalho.

Agradeço à Lisa, pelo imenso apoio nos momentos de frustração, e pelos momentos de descontração e alegria que partilhámos. Agradeço também pelas palavras motivadoras, e que muito contribuíram para a conclusão deste trabalho.

Um agradecimento ao Hugo, pelo grande amigo que foi, e continua a ser. Agradeço também aos meus amigos, por todo o apoio, pelos momentos de alegria, pelas conversas profundas dos almoços e pelas jantaradas alegres no 50.

À minha família, um agradecimento especial, não só pela roupa lavada e comida no prato, mas também pela formação enquanto pessoa, educação e por apostarem fortemente no meu futuro.

Finalmente gostaria de agradecer a todos aqueles que, por alguma razão, não estejam aqui mencionados.

Obrigado

Resumo

A vantagem principal do *design* de fármacos baseado em fragmentos químicos (FBDD) em relação à abordagem convencional, que utiliza moléculas com tamanho de fármacos, é o menor espaço químico a ser explorado. No entanto, devido às baixas afinidades dos fragmentos, os ensaios devem ser feitos através de métodos biofísicos demorados e caros. Para responder a estes problemas, os ensaios poderão ser feitos utilizando metodologias computacionais, que possuem um baixo custo de aplicação e um débito muito elevado.

O *docking* é bastante utilizado em *design* de fármacos com base na estrutura, e apresentou bons resultados em algumas campanhas de FBDD. Apesar das aplicações conseguidas, a maior parte dos estudos foi feita para um conjunto limitado de alvos, sendo que, devido à diversidade de alvos, um estudo sistemático é necessário.

Neste estudo discutimos três questões importantes sobre a aplicação de metodologias de *docking* em FBDD: 1) Será uma biblioteca de fragmentos capaz de amostrar bem as regiões de ligação chave de alvos diferentes? 2) Com que nível de precisão é o *docking* capaz de prever conformações e afinidades de fragmentos em comparação com epítomos de ligandos experimentais conhecidos? 3) Qual o nível de promiscuidade de cada fragmento em relação a uma família variada de alvos? Estas questões foram estudadas através do *docking* de uma biblioteca de 5930 fragmentos contra um conjunto de 32 alvos. Mostramos que as bibliotecas de fragmentos podem amostrar com precisão mais do que uma região chave de ligação, e podem ser utilizadas para dar prioridade a regiões chave em estudos de FBDD. Demonstramos que o *docking* de fragmentos contra epítomos de ligandos conhecidos tem um desempenho de 48,8%. Mostramos, ainda, que alguns fragmentos, independentemente do seu tamanho, serão mais promíscuos para a maior parte dos alvos.

Abstract

The key advantage of fragment-based drug design (FBDD) over conventional drug design, using drug-like molecules, is the much lower chemical space that has to be explored. However, because of the low affinities of fragment molecules, it is necessary to use time consuming and expensive biophysical screening methods. Addressing this issue are computational methodologies, which have a very high throughput and a low cost of application.

Molecular *docking* has widespread use in structure-based design, and has revealed to be reliable in some FBDD campaigns. Despite the successful applications, there is the issue that most studies are directed to a limited set of targets, and, due to the heterogeneous nature of protein targets, a systematic study is necessary.

In this study, we discuss in our work three important questions concerning the application of *docking* methodologies in FBDD: 1) How well a given fragment library is able to sample the binding site hot spots of different targets? 2) How accurate are the *docking* posing and scoring predictions of fragments compared to experimentally known ligand epitopes? 3) What is the degree of promiscuity of each fragment against a diverse family of targets? These questions were studied by *docking* a library of 5930 fragments, against a set of 32 diverse and high-quality targets. We show that fragment libraries can accurately sample more than one binding site hot spot, and can be used to prioritize hot spots in fragment-to-lead studies. We demonstrate that fragment *docking* against known ligand epitopes has a performance of 48.8%. Furthermore, we show that some fragments, independently of their size, are likely to be promiscuous for most targets.

Índice Geral

Agradecimentos	iii
Resumo	v
Abstract	vii
Índice Geral	ix
Abreviaturas, Siglas e Símbolos	xi
Índice de Figuras	xiii
Índice de Tabelas.....	xvii
1. Introdução.....	21
1.1. Desenvolvimento de um fármaco – <i>drug design</i>	21
1.2. Desenho de fármacos baseado na estrutura do alvo	22
Termodinâmica do reconhecimento molecular.....	23
Interações alvo-ligando	26
Regras de Lipinski	32
1.3. Metodologias computacionais no <i>design</i> de novos de fármacos	33
<i>Docking</i> molecular	34
1.4. <i>Design</i> de fármacos baseado em fragmentos químicos.....	40
Limitações da abordagem convencional e espaço químico de fragmentos	40
Energia livre de ligação <i>versus</i> eficiência do ligando.....	42
Do fragmento ao candidato.....	44
Metodologias computacionais em FBDD.....	47
2. Métodos.....	51
2.1. Construção da biblioteca de fragmentos.....	51

2.2.	Seleção de alvos	52
2.3.	<i>Docking</i>	58
2.4.	Análises.....	59
3.	Resultados e discussão	63
3.1.	Amostragem de regiões chave do local de ligação	63
3.2.	RMSD dos fragmentos contra epítomos de ligandos.....	69
3.3.	Classificação de fragmentos	74
3.4.	Especificidade de fragmentos	78
4.	Conclusões	83
5.	Bibliografia.....	87
6.	Anexos.....	101

Abreviaturas, Siglas e Símbolos

DNA – ácido desoxirribonucleico

FBDD –*design* de fármacos baseado em fragmentos

FDA – Food and Drug Administration

FEP – *free energy perturbation*

FQ – *fit quality*

HAC – número de átomos diferentes de hidrogénio

LE – eficiência do ligando

NMR – ressonância magnética nuclear

PDB – *Protein Data Bank*

QSAR – relações de estrutura-atividade quantitativas

RESP –*restrained electrostatic potential*

RMSD – raiz da média dos quadrados dos desvios

RNA – ácido ribonucleico

SAR – relações de estrutura-atividade

SPR – *surface plasmon resonance*

TI – integração termodinâmica

Índice de Figuras

- Figura 1.** Representação da função de Lennard-Jones 12-6. U_{ij} representa a energia potencial entre os átomos i e j , σ é o diâmetro de colisão e ε é o mínimo de energia potencial.38
- Figura 2.** Representação esquemática de um ligando com características de fármaco com afinidade relativamente elevada para o alvo, mas com baixa LE. A afinidade deste tipo de moléculas deriva da superfície molecular em contacto com o recetor, sem que se formem interações de elevada eficiência com as regiões chave do local de ligação. Este tipo de moléculas, na ausência de informação estrutural (num estudo de SAR, por exemplo), torna-se muito difícil de otimizar. As Figuras 4 e 5 mostram o exemplo contrário, numa abordagem baseada em fragmentos químicos.43
- Figura 3.** Representação esquemática da estratégia de evolução de fragmentos. a) o fragmento 1 liga-se uma região chave do local de ligação alvo; b) o fragmento 1 é evoluído através da adição de grupos substituintes de forma a estabelecer interações favoráveis com a superfície superior e a segunda região chave do local de ligação.45
- Figura 4.** Representação esquemática da estratégia de ligação de fragmentos. a) o fragmento 1 liga-se na primeira região chave do local de ligação; b) o fragmento 2 liga-se numa segunda região chave, adjacente à região 1; c) os 2 fragmentos são ligados entre si através de um *linker*, sem que a conformação dos fragmentos 1 e 2 seja alterada.46
- Figura 5.** Núcleos e cadeias laterais utilizadas para a construção da biblioteca de fragmentos. Cada fragmento resulta da combinação de uma núcleo e uma cadeia lateral. Estas estruturas são as mesmas utilizadas por Hartshorn e colaboradores, e são estruturas frequentemente encontradas em fármacos. a) seleção de núcleos utilizados na construção da biblioteca de fragmentos; b) cadeias

laterais para a substituição de átomos de carbono; c) cadeias laterais para a substituição de átomos de nitrogénio. 52

Figura 6. Representação esquemática das regiões chave do local de ligação consideradas. Cada região chave é definida pelo epítipo do ligando experimental que a ocupa. O cálculo de RMSD das soluções de *docking* de fragmentos foi feito contra cada um dos epítipos, separadamente. Valores de RMSD baixos significarão que as respetivas soluções estarão próximas da região chave ocupada pelo epítipo correspondente. 64

Figura 7. Número de soluções de *docking* de fragmentos a diferentes intervalos de RMSD em relação a um dado epítipo de ligando experimental, para 5 alvos selecionados. Ligandos com mais do que um epítipo estão legendados. a) alvos 1TZ8 e 1HWI; b) epítipos A e B do alvo 1KE5; c) epítipos A e B do alvo 1YGC e d) epítipos A, B e C do alvo 1LPZ..... 65

Figura 8. Localização do centro geométrico, representado em pequenas esferas amarelas, de todos as soluções de *docking* de fragmentos no local de ligação de: a) 1TZ8; b) 1HWI; c) 1KE5; d) 1YGC e e) 1LPZ. Os ligandos experimentais estão representados *sticks* com coloração CPK e cada epítipo está representado na forma de *ball-and-stick* com um esquema de cores diferente (no qual as cores dos carbonos variam entre ciano, magenta, laranja e rosa). A proteína está representada em *ribbons* de cor verde e uma superfície molecular com o seguinte esquema de cores: nitrogénio – azul, carbono – verde, oxigénio – vermelho. Os epítipos e as regiões chave do local de ligação correspondentes estão assinalados como A, B e C. Foi aplicado um corte transversal da proteína para facilitar a visualização do local de ligação. 67

Figura 9. RMSD entre o epítipo experimental e o fragmento da biblioteca equivalente. Os alvos cujos ligandos com mais do que um

epítopo estão assinalados com a, b ou c sobrescrito. No canto superior direito está representado o gráfico da distribuição de valores de RMSD.70

Figura 10. Conformações de ligação de ligandos experimentais dos alvos 1R90, 1X8X, 1OWE, 1LPZ-A, 1LPZ-B, 1LPZ-C, 1J3J-A, 1J3J-B e os fragmentos da biblioteca equivalentes a cada um destes. Os ligandos experimentais estão representados com *sticks* com o seguinte esquema de cores: nitrogénio – azul, carbono – magenta, oxigénio – vermelho e hidrogénio – branco. Os fragmentos estão representados na forma *ball-and-stick*. A estrutura secundária da proteína está representada como *cartoon* verde e os resíduos chave estão representados como *sticks* com o código de três letras correspondente e o número de resíduo. Os epítopos e as regiões chave do local de ligação estão assinaladas com A, B e C. As interações chave entre os ligandos e a proteína estão representados por linhas tracejadas amarelas.71

Figura 11. Conformação dos ligandos experimentais da β -secretase, com códigos PDB a) 2OHL e b) 2OHM, e a os fragmentos equivalentes da biblioteca com estados de protonação alternativos. Os ligandos experimentais estão representados como *sticks* com o código de cores seguinte: nitrogénio – azul, carbono – magenta, oxigénio – vermelho, hidrogénio – branco. Os fragmentos estão representados como *ball-and-stick*. A estrutura 2D dos diferentes estados de protonação considerados para cada fragmento está representada no canto superior esquerdo de cada painel. A estrutura secundária do alvo está representada na forma de *cartoon* verde, com os resíduos chave em *sticks* e o correspondente código de três letras e número. As interações chave entre o ligando e o alvo estão representadas como linhas tracejadas amarelas.73

Figura 12. RMSD em relação ao epítopo em função de a) *fit quality* (FQ); b) eficiência do ligando (LE) para o alvo 1OWE. As soluções de *docking* de fragmentos com maior FQ e LE estão assinaladas com um asterisco (*) e a estrutura dos complexos está representada na Figura 13a e b....75

Figura 13. Conformações de ligação do ligando experimental do alvo 1OWE e de dois fragmentos com o FQ mais elevado obtido por *docking* (painéis a e b). O painel c representa a versão aromática do fragmento em a.....76

Figura 14. Média das medidas de *scoring* utilizadas em função do número de átomos diferentes de hidrogénio. Nos painéis a e b a medida é a LE para: a) todos os alvos; b) o alvo 1HWI. Nos painéis c e d a medida é o FQ para: c) todos os alvos; d) o alvo 1HWI. Nos painéis b e d são representados os intervalos de confiança de 95% para cada uma das médias apresentadas.....77

Figura 15. Posição média na classificação pela medida de FQ contra todos os alvos vs. desvio padrão da média. Os dados foram divididos em 4 intervalos de HAC: a) 6-9; b) 10-13; c) 14-17; d) 18-21.....79

Índice de Tabelas

Tabela 1. Distâncias típicas das interações não covalentes específicas mais frequentemente encontradas na análise feita por Bissantz et al.⁹.....27

Tabela 2. Alvos selecionados para as experiências de *docking* da coleção Astex Diverse Set⁹⁹. Estão também representados os ligandos respectivos a cada alvo e os epítomos considerados destacados em caixas azuis para epítomos A, vermelhas para epítomos B e verdes para epítomos C.53

Introdução

1. Introdução

Nas últimas duas décadas, a utilização de metodologias computacionais tem tido uma crescente relevância no desenvolvimento de novos fármacos. O baixo custo e o elevado débito de informação são os principais fatores que tornam estas metodologias atraentes, tanto em aplicações industriais como acadêmicas. Embora não exista nenhum fármaco desenvolvido utilizando apenas métodos computacionais, a sua utilização abrange uma grande parte das etapas do desenvolvimento de um fármaco, desde a identificação de uma estrutura química base para uma série de candidatos, passando pelo teste de milhares de compostos virtuais contra um o alvo terapêutico, até às etapas mais avançadas de otimização.

1.1. Desenvolvimento de um fármaco – *drug design*

O desenvolvimento de um fármaco dirigido a uma determinada patologia passa por diversas e exaustivas etapas, que envolvem equipas multidisciplinares. A fase inicial passa pela identificação do alvo terapêutico, que geralmente é uma proteína, podendo também ser DNA ou RNA. No caso de uma campanha bem-sucedida, o fármaco é aprovado pelas entidades governamentais. Entre a identificação do alvo e os ensaios clínicos está o desenvolvimento do candidato que geralmente representa uma família de moléculas similares estruturalmente, e que vai sendo modificado de acordo com a informação obtida através dos ensaios efetuados no decorrer da campanha. Este processo de criação, modificação e avaliação do candidato é o chamado *drug design*, ou desenho de fármacos.

Existem, essencialmente, duas abordagens ao desenho de fármacos: o desenho baseado na estrutura de ligandos, e desenho baseado na estrutura do alvo. Estas duas abordagens, como o seu nome diz, distinguem-se entre si pela referência estrutural utilizada para construir a nova molécula.

No desenho baseado na estrutura de ligandos, é utilizado um ou mais ligandos já conhecidos e com atividade para o alvo terapêutico em estudo. A nova molécula é construída a partir da informação estrutural desse conjunto de ligandos conhecidos. Esta abordagem parte do princípio chamado de relações estrutura-atividade, ou SAR, em que se assume que moléculas similares, química e estruturalmente, terão atividades semelhantes. Deste modo, um químico medicinal irá testar, numa estratégia de tentativa e erro, quais os grupos químicos importantes para a atividade do ligando contra o alvo terapêutico, e que grupos e/ou posições novas poderão ser exploradas, sendo que deste processo resulta um conjunto de parâmetros estéreos e eletrónicos chamado de farmacóforo¹. Outra metodologia dentro deste tipo de abordagem ao desenho de fármacos é chamada de relação quantitativa de estrutura-atividade, ou QSAR¹. Esta metodologia utiliza modelos obtidos por regressão, a partir das características físico-químicas (lipofilicidade, polarizabilidade, densidade eletrónica, efeitos estéreos, etc.) e valores atividade de ligandos conhecidos para um determinado alvo, e assim inferir sobre as características necessárias ao novo ligando para que este possua atividade ótima. As desvantagens da utilização deste tipo de abordagem assentam principalmente no facto de nem todas as moléculas estruturalmente similares possuírem atividades semelhantes. A limitação da informação estrutural do alvo terapêutico apenas a interações conhecidas ou exploradas pelos ligandos existentes e, por consequência, a dificuldade em encontrar novas interações alvo-ligando são outras desvantagens associadas a este tipo de abordagem.

1.2. Desenho de fármacos baseado na estrutura do alvo

No desenho de fármacos com base na estrutura do alvo terapêutico a informação estrutural deste último é conhecida através da sua estrutura de cristalografia de raios-X ou de ressonância magnética nuclear (NMR). Isto permite que o ligando seja construído à

medida do alvo terapêutico, explorando as interações descritas na estrutura deste. Os novos ligandos são testados contra a estrutura do alvo, sendo o modo de ligação estudado através da co-cristalização do ligando e do alvo terapêutico, ou das alterações no espectro de NMR²⁻⁴.

A principal vantagem deste tipo de abordagem é que, ao contrário da abordagem baseada no ligando, a informação sobre o modo de ligação entre o ligando e o alvo é obtida diretamente através da estrutura de cristalografia de raios-X, ou espectro de NMR, do complexo alvo-ligando, permitindo saber quais as interações favoráveis ou desfavoráveis do ponto de vista entálpico. A informação estrutural do alvo permite, ainda, saber quais as novas interações que podem ser exploradas sem recorrer a métodos exaustivos de tentativa e erro. Deste modo, não será necessário a criação de novas relações estrutura-atividade para encontrar novas interações, apenas a introdução de grupos substituintes em posições que permitam que estas se estabeleçam.

No entanto, nem todos os alvos, ou complexos alvo-ligando, são facilmente cristalizáveis. Outra questão reside no facto de a estrutura de cristalografia de raios-X representar apenas uma média de todas as conformações que o ligando e o alvo podem ter, e de a ressonância magnética nuclear apenas permitir obter a estrutura de alvos de pequenas dimensões. Estas questões são algumas desvantagens da utilização experimental do *design* de fragmentos com base na estrutura do alvo.

Termodinâmica do reconhecimento molecular

No processo de reconhecimento molecular, tal como em qualquer outro processo espontâneo, é necessário que a sua energia livre (ΔG) tenha um valor negativo, ou seja, que a soma das componentes entálpica (ΔH) e entrópica ($-T\Delta S$) seja inferior a zero. A componente entálpica envolve, de uma forma resumida, o balanço das interações

favoráveis e desfavoráveis à formação do complexo alvo-ligando, e das energias de solvatação e dessolvatação do ligando e do local de ligação do alvo. A componente entrópica envolve todos os movimentos rotacionais, translacionais e conformacionais de todas as moléculas do sistema em estudo, isto é, do ligando, do alvo e do solvente. Como regra geral, parece existir um mecanismo de compensação entre as componentes entálpica e entrópica na ligação entre um ligando e o seu recetor^{5, 6}. Isto significa que, o quanto mais direcionada ou forte for uma interação, maior rigidez esta estará a impor às moléculas envolvidas nesta, ou seja, quanto mais entalpicamente favorável for a interação, mais desfavorável do ponto de vista entrópico será. No entanto, a natureza destes mecanismos é muito dependente do sistema em estudo. Existem, por exemplo, casos em que a energia livre resultante de várias interações não covalentes é superior, em módulo, à soma das suas energias livres individuais, isto é, são interações positivamente cooperativas. Um exemplo de cooperatividade positiva é o caso da ligação da biotina à streptavidina, uma das interações proteína-ligando mais fortes observadas na natureza, no qual foi observado um melhor empacotamento da streptavidina após ligação da biotina⁷. Este empacotamento faz com que a distância de algumas interações fique mais curta, tornando-as mais favoráveis do ponto de vista entálpico. Esta cooperatividade faz com que o ganho de entalpia seja muito elevado em relação à perda de entropia causada pela maior rigidez do complexo resultante.

Este efeito de cooperatividade já foi observado noutros sistemas, nomeadamente em complexos proteína-proteína⁸, e o caso do complexo streptavidina-biotina está largamente estudado (observável pelas 136 estruturas de raios-X depositadas até à data no PDB). No entanto, explorar o efeito de cooperatividade positiva na criação de novos fármacos é muito difícil.

Existem grandes barreiras no desenho de candidatos com boas afinidades para o seu alvo terapêutico, usando uma abordagem baseada na estrutura. A heterogeneidade dos locais de ligação faz com que mecanismos empíricos, desenvolvidos pelos químicos medicinais ao longo da sua experiência, tenham um espectro de aplicabilidade reduzido. Existe também uma visão tendencialmente “entálpica” e rígida dos complexos ligando-alvo, isto é, focada nas interações formadas entre o ligando e o alvo, e que dá menor atenção à flexibilidade que afeta a componente entrópica. Outros efeitos que tendem a ser ignorados estão relacionados com a solvatação do local de ligação e do ligando, que afeta consideravelmente o perfil termodinâmico do processo de reconhecimento molecular⁹.

Um dos efeitos mais importantes no processo de reconhecimento molecular é o efeito hidrofóbico. O conceito clássico deste efeito é descrito da seguinte forma: uma molécula apolar é rodeada por moléculas de água, bastante ordenadas e ligadas entre si por ligações de hidrogénio, causando uma diminuição de entropia. Como esta diminuição é tanto maior quanto mais superfície das moléculas apolares estiver exposta, estas moléculas acabam por agregar-se e, assim, diminuir a superfície exposta a água, resultando numa menor diminuição de entropia. O mesmo conceito se pode aplicar à ligação entre uma proteína e ligando apolares. A incorporação do ligando no local de ligação da proteína diminui a superfície apolar exposta ao solvente, aumentando a entropia. No entanto verifica-se que a ligação entre uma proteína e ligando apolares é, geralmente, um processo dominado pela componente entálpica, ou seja, é conduzida pelo efeito hidrófobo “não clássico”. A solvatação incompleta de locais de ligação lipofílicos baixa os custos entálpicos associados favorecendo a formação do complexo proteína ligando¹⁰⁻¹². Também foi observado que, no caso de locais de ligação hidrofílicos e completamente solvatados, ligandos, que aparentemente “encaixariam” nesse local, apresentam pouca ou nenhuma afinidade devido aos elevados custos

de solvatação¹³. Assim, torna-se importante para o químico medicinal, numa campanha de *design* de fármacos com base na estrutura do alvo, ter em conta estes mecanismos e estudar a solvatação do local de ligação da proteína não complexada, de modo a evitar os custos para a entalpia, e, conseqüentemente, na energia livre de ligação.

Interações alvo-ligando

O reconhecimento molecular de um fármaco e o seu alvo terapêutico depende essencialmente da existência de interações favoráveis específicas entre as duas moléculas. Estas interações favoráveis podem ser ligações covalentes ou interações intermoleculares não covalentes.

Os fármacos que interagem covalentemente com o alvo terapêutico alteram de forma irreversível a sua função. Este tipo de fármacos, sendo o caso mais conhecido o ácido acetilsalicílico, ou aspirina¹⁴, é muito reduzido em número, em muito devido à dificuldade de criar moléculas reativas específicas para um determinado alvo, e de, em grande parte dos casos, não ser clinicamente favorável a alteração irreversível do alvo terapêutico e, conseqüentemente, da via metabólica em que este está envolvido.

As interações intermoleculares não covalentes são as mais exploradas no processo de desenho de novos fármacos. Estas interações são mais fracas que as covalentes, e podem ser, por ordem decrescente de força, interações ião-dipolo, ligações de hidrogénio, interações dipolo-dipolo, dipolo-dipolo induzido e de dispersão de London.

Numa análise estatística a uma coleção de estruturas de cristalografia de raios-X de complexos proteína-ligando, Bissantz e seus colaboradores⁹ observaram, com maior frequência, as seguintes interações intermoleculares específicas: ligações de hidrogénio, ligações de halogénio, interações multipolares ortogonais, interações entre halogénios e anéis aromáticos, interações hidrofóbicas nas quais

se incluem as interações entre grupos aromáticos e interações de grupos aromáticos com grupos alquila, e finalmente interações envolvendo o átomo de enxofre.

As ligações, ou pontes, de hidrogénio são as interações específicas mais importantes no processo de reconhecimento molecular. Estas interações atrativas formam-se entre um átomo de hidrogénio de uma molécula, ou fragmento molecular X-H, em que o átomo dador X é mais eletronegativo que o átomo de hidrogénio, e uma molécula ou fragmento com um átomo aceitador Y rico em eletrões¹⁵.

Tabela 1. Distâncias típicas das interações não covalentes específicas mais frequentemente encontradas na análise feita por Bissantz *et al.*⁹

Tipo de interação		Intervalo típico de distâncias	
Ligações de hidrogénio	OH	NH	
O Carbonilo/Sulfonilo	2.7 – 3.0	2.8 – 3.1	
N heteroaromático	2.7 – 3.0	2.8 – 3.1	
O ácido carboxílico	2.6 – 2.8	2.7 – 3.0	
Ligações de halogénio	Cl	Br	I
O carbonilo	3.0 – 3.4	3.0 – 3.5	2.9 – 3.5
Interações multipolares			
F ⋯ C carbonilo	3.0 – 3.7		
Interações com C alifático			
F	3.3 – 3.9		
Cl	3.6 – 4.3		
C alifático	3.7 – 4.4		
O sulfonilo	3.3 – 3.9		
S divalente	3.8 – 4.2		
Interações com C arilo	No plano	Acima do plano	
C aromático ⋯ S divalente	3.7 – 4.2		
C aromático ⋯ F	3.3 – 3.7		
C aromático ⋯ Cl	3.6 – 4.1	3.5 – 4.0	
C aromático ⋯ Br	3.7 – 4.2	3.5 – 4.1	
C aromático ⋯ -CH ₂ -	3.8 – 4.4	3.6 – 4.2	
C aromático ⋯ CH ₃ O-		3.4 – 4.0	
C aromático ⋯ CN ⁺		3.4 – 4.0	
C aromático ⋯ N amida		3.4 – 3.8	
	Paralelos e deslocados	Edge-to-face	
C aromático ⋯ C aromático	3.4 – 3.6	3.6 – 3.8	

Em alvos terapêuticos proteicos, as ligações de hidrogénio mais frequentes têm como aceitadores oxigénios de grupos carbonilo, sulfonilo, éteres e ésteres, e átomos de nitrogénio em heterociclos aromáticos. As distâncias médias entre os vários tipos de aceitadores de ligações de mais comuns e grupos hidroxilo e amina como doadores estão descritos na Tabela 1.

Para além de a força das ligações de hidrogénio ser altamente dependente da distância entre o dador e o aceitador, estas também possuem uma "preferência angular" bastante pronunciada. Geralmente as ligações de hidrogénio são formadas na direção do par de eletrões livre do aceitador. Um exemplo é o grupo carbonilo cujo ângulo preferido entre o carbono e oxigénio do deste grupo, e o hidrogénio do dador é aproximadamente 120° , que corresponde à direção dos pares de eletrões livres do oxigénio do carbonilo. No entanto esta regra não se aplica a grupos sulfonil, em que a ligação de hidrogénio ao longo do eixo da ligação dupla enxofre oxigénio. Quando o aceitador faz parte de um sistema com eletrões π , o ângulo entre o plano do aceitador – caso seja um carbonilo o plano é formado pelo carbonilo em si e pelos seus substituintes, ao passo que no caso de um anel aromático o plano é formado pelo anel em si – e a ligação de hidrogénio é geralmente abaixo dos $25-30^\circ$ ¹⁶. Um sumário das geometrias preferidas na formação de ligações de hidrogénio está descrito na tabela 1.

As ligações de halogénio ocorrem entre um halogénio ligado covalentemente e um grupo com um ou mais pares de eletrões livres, por exemplo uma base de Lewis. Os halogénios envolvidos neste tipo de ligações são o cloro, bromo e iodo. Embora os halogénios ligados covalentemente seja vistos como tendo um caráter eletronegativo, dificultando a compreensão de como dois átomos com caráter eletronegativo se podem atrair, Brinck e seus colaboradores¹⁷ e, mais tarde, outros^{18, 19} mostraram que halogénios em algumas moléculas possuem regiões de potencial eletrostático positivo. Mais tarde Clarke e

seu grupo²⁰ chamaram a essa região de σ -hole (ou buraco σ). A explicação proposta para a existência do σ -hole tem a ver com o facto de os três pares de eletrões livres do halogénio terem uma configuração aproximada $s^2p^2p^2$, formando zona de carga negativa em torno da região central da sua região central, enquanto a terceira orbital p ao longo do eixo carbono-halogénio está distorcida em direção ao carbono formando a ligação covalente simples carbono-halogénio. As ligações de halogénio são significativamente mais fracas que as ligações de hidrogénio²¹. No entanto estas interações estão associadas a grandes aumentos na afinidade, observados em várias campanhas de *design* de inibidores²²⁻²⁵, e que se devem, em grande parte, ao menor custo em termos de dessolvatação do que as ligações de hidrogénio.

As interações multipolares ortogonais são caracterizadas por contactos próximos entre dois grupos funcionais dipolares ortogonais entre si²⁶. Estas interações são frequentes envolvendo substituintes contendo flúor, sendo encontrados a distâncias curtas de carbonos de grupos carbonilo, com a ligação covalente C-F ortogonal ao plano da ligação C=O⁹. Esta interação flúor-carbonilo é fraca embora vários exemplos de SAR revelem aumentos na afinidade de ligação²⁷. Outros halogénios também podem participar em interações multipolares ortogonais, mas menos frequentemente do que no caso do flúor. Substituintes contendo cloro preferem uma disposição paralela ao plano do grupo carbonilo, em relação à disposição ortogonal, sendo a ligação de halogénio a interação mais frequente envolvendo o cloro. A frequência de interações multipolares, ortogonais ou não, diminui ao longo do grupo dos halogénios, aumentando a frequência de ligações de halogénio⁹.

Outro tipo de interações intermoleculares específicas envolvendo halogénios ocorrem com anéis aromáticos. Geralmente a interação faz-se entre um carbono do anel e o halogénio do substituinte, geralmente com os eixos C-H e C-X formando um ângulo superior a 120° e com

distâncias entre o halogénio e hidrogénio abaixo do raio de van der Waals⁹. Desiraju classificou esta interação como ligação de hidrogénio “muito fraca” ao observar, num estudo cristalográfico, contactos entre flúor e hidrogénio abaixo do raio de van der Waals ($\sim 2.6 \text{ \AA}$) à medida que o conteúdo em flúor do grupo arilo aumentava, isto é, com o aumento da acidez do hidrogénio²⁸. No caso de halogénios suscetíveis de possuírem um σ -hole é possível encontrar, com menor frequência, pontes de halogénio com os sistemas π do anel aromático, numa disposição ortogonal ao plano formado por este⁹. No entanto, embora seja uma interação atrativa^{29, 30}, esta não é suficientemente alta para justificar alguns ganhos de afinidade associados à introdução de substituintes halogenados.

As interações hidrofóbicas são as que permitem, geralmente, os maiores ganhos de afinidade de ligação em química medicinal. Há vários estudos que mostram que o melhor parâmetro estrutural correlacionado com a afinidade de ligação é a quantidade de superfície hidrofóbica “enterrada” aquando da formação de um complexo proteína-ligando³¹. Uma regra empírica estabelecida para química *host-guest*, por Mecozzi e seus colegas diz que a ligação é ótima quando 55% do volume do hóspede é preenchido³². Esta regra também é aplicável em química medicinal, como foi confirmado por Zucher *et al.*³³. Assim, enormes ganhos em afinidade de ligação podem ser obtidos com o preenchimento ótimo de bolsas hidrofóbicas com átomos apolares de um ligando³⁴⁻³⁶. Estes ganhos de afinidade estão associados, em grande parte, à baixa solvatação das bolsas hidrofóbicas. Estas bolsas estão ocupadas por um número reduzido de moléculas de água, que estabelecem poucas interações^{37, 38}. Isto leva a uma redução dos custos de dessolvatação e a um aumento da flexibilidade molecular residual que, devido à ocupação parcial da bolsa e a natureza menos forte das interações hidrofóbicas, reduz a penalização entrópica da formação do complexo proteína-ligando.

Dentro das interações hidrofóbicas, as interações envolvendo aminoácidos com anéis aromáticos, como a histidina, fenilalanina, tirosina e triptofano, são bastante comuns no contexto de complexos proteína-ligando³⁹⁻⁴¹. Interações entre os sistemas π de dois anéis aromáticos podem seguir, predominantemente, duas orientações isoenergéticas – em forma de T em que os planos dos anéis são perpendiculares entre si, chamada de *edge-to-face*, e num arranjo com os planos dos anéis empilhados e paralelos entre si, e deslocados horizontalmente, chamada de empilhamento π - π . Destas duas orientações, a última é a mais frequente entre os aminoácidos das proteínas⁴², sendo que a orientação é bastante dependente dos substituintes dos e da inserção de heteroátomos nos anéis aromáticos. Geralmente a orientação empilhada é mais forte quando ambos os anéis são deficientes em eletrões, ou seja, possuem substituintes eletro-retiradores, ou quando um deles é deficiente e o outro é rico em eletrões, isto é, quando o último possui substituintes dadores de densidade eletrónica. Quando um dos hidrogénios é mais ácido, a preferência pela orientação *edge-to-face* aumenta. Este aumento de acidez pode ser proporcionado pela presença de um grupo eletro-retirador em posição orto- e/ou para- do anel aromático⁹.

A orientação *edge-to-face* não se aplica só entre grupos anilo, podendo ocorrer entre um anel aromático e um grupo alquilo, em que um átomo de hidrogénio do grupo alquilo interage com o sistema π do anel. A força desta interação aumenta quanto mais ácido for o átomo de hidrogénio^{43, 44}. É possível encontrar também interações π - π e *edge-to-face* envolvendo anéis aromáticos e redes de ligações de hidrogénio como pares guanidínio-carboxilato ou com os sistemas π de ligações amida⁹.

Os sistemas π de anéis aromáticos podem também fazer interações relativamente fortes com catiões. Um exemplo conhecido é o da carbamoylcolina, que se liga ao recetor da acetilcolina através de interações entre os grupos amina primários protonados e os anéis

aromáticos dos resíduos de tirosina e triptofano. A nicotina também se liga a este recetor de uma forma similar, variando nos resíduos aromáticos com que é feita a interação catião- π ⁴⁵.

A habilidade do átomo de enxofre divalente interagir com grupos funcionais pobres ou ricos em eletrões é um indicador da sua versatilidade. Grupos ricos em eletrões têm tendência a aproximarem o átomo de enxofre ao longo da extensão da ligação enxofre-carbono, na direção das orbitais σ antiligantes (σ^*), sendo a interação com o grupo carbonilo um exemplo. Os grupos pobres em eletrões têm tendência a interagir com os pares de eletrões livres⁴⁶. Interações metionina-arilo são bastante frequentes em estruturas proteicas, sendo as interações com o carbono aromático as mais observadas envolvendo o enxofre da metionina (cerca de 40% das interações)⁴⁷, enquanto a interação com o sistema π é apenas observada em cerca de 22% dos casos⁹. Esta última é observada frequentemente em locais de ligação de anéis de adenina, como é o caso da catecol-O-metiltransferase com o seu cofator S-adenosilmetionina, em que o fragmento C-S-C da metionina está posicionado de uma forma coplanar, acima do anel de purina à distância de van der Waals ($\sim 4\text{\AA}$)⁴⁸.

Um resumo das distâncias de todos os tipos de interações referidos anteriormente está descrito na Tabela 1.

Regras de Lipinski

A termodinâmica do reconhecimento molecular e as interações ligando-alvo constituem apenas a parte farmacodinâmica do *design* de novos fármacos – o facto de uma molécula ligar-se com elevada afinidade a um alvo terapêutico, por si só, não significa que esta será ativa quando administrada. Existem outras propriedades que se devem ter em conta e que fazem parte da farmacocinética de um fármaco, isto é, a absorção, distribuição, metabolização e excreção. Propriedades como a estabilidade, a reatividade, solubilidade e

permeabilidade são exemplos de propriedades que afetam a farmacocinética. Assim, a incorporação destes conceitos no início das campanhas de *design* de novos fármacos é essencial.

As regras de Lipinski, ou “regra dos cinco” é um conjunto de condições empíricas inferidas por Lipinski e Lombardo⁴⁹ num trabalho de análise estatística de propriedades de fármacos conhecidos. Segundo Lipinski, um fármaco ativo via administração oral não deve violar mais do que uma das seguintes condições:

1. não mais do que cinco doadores de ligação de hidrogénio, contabilizando a soma de grupos hidroxilo e amina;
2. massa molecular até 500 inclusive;
3. coeficiente de partição octanol-água abaixo de 5;
4. não mais do que 10 aceitadores de ligação de hidrogénio, contabilizando a soma de oxigénios e nitrogénios aceitadores.

Uma importante adição a este conjunto de regras foi feita por Veber e seus colaboradores⁵⁰, que limita a um máximo de 10 o número de ligações capazes de rodar.

Moléculas que façam parte de classes de compostos que funcionam como substratos para transportadores biológicos são consideradas exceções para esta regra. Na sua análise, Lipinski *et al.* as exceções encontradas eram vitaminas, antibióticos, antifúngicos e glicosídeos cardíacos⁴⁹.

1.3. Metodologias computacionais no *design* de novos fármacos

Embora não exista nenhum caso de um fármaco desenhado exclusivamente através de métodos computacionais, a aplicação destes métodos em campanhas de *design* de novos fármacos é cada vez mais comum. O baixo custo de aplicação e a maior e crescente rapidez permitem uma redução significativa da carga experimental e da quantidade de materiais e reagentes.

As metodologias computacionais podem ser aplicadas em ambos os tipos de *design* de fármacos referidos acima – *design* com base em ligandos conhecidos e *design* com base na estrutura do alvo. Quando a estrutura do alvo não é conhecida, pode ser aplicado o QSAR computacional, em que um conjunto de ligandos com atividade conhecida é analisada para a criação do farmacóforo, e uma biblioteca de moléculas conhecidas ou propostas é testada para encontrar potenciais candidatos⁵¹.

Quando a estrutura do alvo é conhecida, metodologias como o 3D-QSAR, o *docking* molecular e o FEP-TI são as mais comuns.

O 3D-QSAR segue o mesmo princípio do QSAR convencional (a duas dimensões) mas utiliza a estrutura tridimensional do alvo e do ligando complexados para obter os parâmetros do modelo de ajustamento⁵².

A técnica de FEP-TI, ou integração termodinâmica, consiste na transformação alquímica gradual de um ligando de referência e de energia livre de ligação conhecida, numa nova molécula de estrutura relativamente similar. O cálculo da energia livre da transformação alquímica de uma molécula em outra semelhante permite obter a diferença de energias livres de ligação entre o ligando inicial e o final⁵³.

Docking molecular

Descrito pela primeira vez por Kuntz e seus colaboradores⁵⁴, o *docking* molecular é a metodologia computacional mais utilizada no *design* de fármacos com base na estrutura do alvo. O *docking* consiste em prever a conformação e orientação de um ligando num local de ligação alvo, através da utilização de duas funções diferentes – o algoritmo de procura e a função de *scoring*.

Algoritmos de procura

O algoritmo de procura é, de uma forma simplificada, a função que, como o nome diz, faz a procura das conformações do ligando que melhor “encaixam” no centro ativo do recetor. Existem três tipos de

algoritmos de procura, que se distinguem pela forma como a flexibilidade do ligando é tratada⁵⁵ – métodos de procura sistemática, métodos aleatórios e métodos de simulação.

Métodos de procura sistemática. Métodos que tentam explorar todas as conformações possíveis do ligando são chamados de métodos de procura sistemática. No entanto estes métodos enfrentam o problema do aumento exponencial de combinações com o aumento dos graus de liberdade do ligando⁵⁶. Uma forma de ultrapassar este problema é fazer um crescimento incremental do ligando no local de ligação. Esta procura incremental pode ser feita através do *docking* de pequenos fragmentos molecular e liga-los covalentemente entre si. Outra forma é dividir os ligandos em estruturas rígidas – fragmentos base – e estruturas flexíveis – cadeias laterais. Neste último caso, depois de definidas as estruturas rígidas, é feito o *docking* destas, e as cadeias laterais são adicionadas passo a passo⁵⁷⁻⁵⁹. Uma alternativa à procura incremental é a utilização de bibliotecas de conformações moleculares, previamente geradas, para um ligando, as quais são utilizadas no *docking* como moléculas rígidas⁶⁰.

Métodos de procura aleatória. Estes métodos introduzem alterações aleatórias na rotação de ligações e translação/rotação do ligando e avaliam a nova conformação obtida usando uma função de probabilidade. Os algoritmos mais populares para este tipo de procura são algoritmos genéticos e de Monte Carlo.

Os algoritmos de Monte Carlo geram conformações, rotação e translação iniciais aleatórias no local de ligação e avalia o “encaixe” do ligando. Seguidamente, uma nova conformação é gerada de forma aleatória e avaliada. Se esta última tiver maior afinidade do que a conformação do passo anterior, a nova conformação é aceite. Caso contrário, esta é rejeitada dentro de uma certa probabilidade. Este processo de criação e avaliação de novas conformações é repetido até que o número de conformações desejadas seja obtido⁵⁶.

Algoritmos genéticos são baseados nos conceitos de genética e evolução biológica. A população, neste caso o ligando, é descrita por um conjunto de valores que descrevem a conformação, rotação e translação do ligando – as variáveis de estado – sendo que cada variável de estado corresponde a um gene. O estado do ligando, definido pelas variáveis de estado, corresponde ao genótipo, e as coordenadas de cada átomo do ligando, que traduzem a conformação e orientação deste, correspondem ao fenótipo. Pares aleatórios de indivíduos (ligandos com estados diferentes) são “cruzados” usando um processo de *crossover*, em que os novos indivíduos são herdam genes dos “progenitores”. Um processo que pode ocorrer é a “mutação” aleatória, no qual um gene pode ser alterado num valor aleatório. Os novos estados são avaliados pela função de scoring, que atua como meio de “seleção natural”. Assim as soluções que melhor preenchem o local de ligação do recetor irão reproduzir-se enquanto as restantes são rejeitadas⁶¹.

Métodos de Simulação. Ao contrário dos métodos estocásticos ou aleatórios que introduzem alterações aleatórias nos ligandos, métodos de simulação utilizam as equações, como a do movimento de Newton, para prever a próxima posição do ligando na superfície de energia potencial a partir da posição anterior.

Simulações de dinâmica molecular são a técnica mais utilizada, e resolvem a equação do movimento de Newton

$$F_i = m_i a_i$$

em que F é a força, m é a massa e a é a aceleração. Um problema associado à dinâmica molecular é a frequente incapacidade de ultrapassar elevadas barreiras energética durante os tempos comuns de simulação. Isto faz com que, em vez do desejado mínimo global, apenas seja obtido um mínimo local da superfície de energia potencial do ligando⁵⁵. Formas de ultrapassar este problema podem ser simular o sistema proteína-ligando a diferentes temperaturas⁶², ou iniciar as simulações a partir de diferentes estados do ligando.

Um método distinto é a minimização de energia. No entanto esta só é capaz de encontrar mínimos de energia locais, por isso é geralmente acoplada a outros métodos de procura como Monte Carlo⁶³ ou algoritmos genéticos⁶¹.

Funções de scoring

Avaliar e classificar as conformações previstas de um ligando em termos de uma medida da energia livre associada à formação do complexo com o local de ligação do recetor, é essencial, não só para o processo de *docking* em si, mas também para a distinção entre ligandos diferentes, que é um aspeto crucial para o *design* de novos fármacos.

Existem técnicas de simulação bastante rigorosas que calculam energias livres de ligação^{64, 65}, como FEP/TI, mas que são demasiadamente dispendiosas do ponto de vista computacional, e que não são práticas quando é necessário avaliar longas bibliotecas de ligandos.

As funções de *scoring* utilizadas em *docking* fazem um balanço entre precisão e velocidade de cálculo, de forma a permitirem o teste de vários milhares de moléculas contra um determinado recetor em tempo útil. Esse balanço é feito com a introdução de aproximações e simplificações, que fazem com que fenómenos físicos, como a entropia, não sejam totalmente representados. Existem três tipos de funções de *scoring* – baseadas em campos de forças, empíricas e baseadas em conhecimento estrutural.

Funções baseadas em campos de forças. Campos de forças de mecânica molecular geralmente quantificam a soma da energia interna do ligando e a energia das interações proteína-ligando. Na maior parte das funções de *scoring* baseadas em campos de forças, a energia interna da proteína não é incluída devido ao facto de considerarem apenas uma conformação desta, o que simplifica significativamente o *scoring*.

As interações proteína-ligando são, geralmente, descritas por termos energéticos de van der Waals e eletrostáticos⁶⁶. Estes termos são dados por uma função de potencial Lennard-Jones⁵⁶:

nos termos eletrostáticos temos para duas moléculas, A e B

$$E_{\text{coul}}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

em que N é o número de átomos nas moléculas A e B e q é a carga em cada átomo;

nos termos de van der Waals temos para dois átomos, i e j

$$E_{\text{vdW}}(r) = \sum_{j=1}^N \sum_{i=1}^N 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

em que ϵ é o mínimo de energia potencial e σ é o diâmetro de colisão entre os átomos i e j . A equação utilizada é uma função de Lennard-Jones 12-6, em que o termo com expoente 12 é o termo das repulsões de curta-distância e o termo com expoente 6 é o termo das forças atrativas que se aproxima de zero à medida que a distância entre os átomos aumenta⁵⁶. A representação

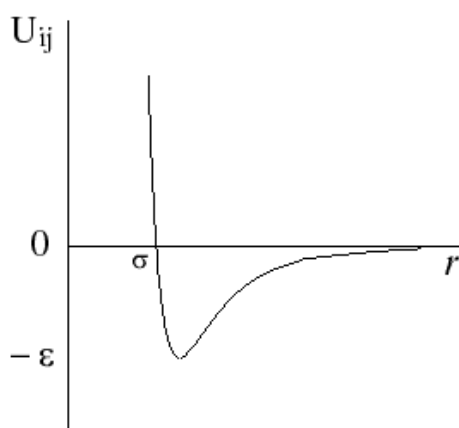


Figura 1. Representação da função de Lennard-Jones 12-6. U_{ij} representa a energia potencial entre os átomos i e j , σ é o diâmetro de colisão e ϵ é o mínimo de energia potencial.

gráfica da função é apresentada na Figura 1.

A função que descreve a energia interna do ligando é bastante similar à da interação proteína-ligando, dado que também inclui termos eletrostáticos e de van der Waals⁶⁶.

Uma limitação deste tipo de funções é a introdução de raios de corte para distâncias de interações não covalentes, o que reduz a precisão do cálculo das contribuições de interações de longa distância que participam no reconhecimento molecular⁶⁶.

Funções de *scoring* empíricas. Este tipo de funções tenta reproduzir as energias de ligação e/ou conformações de ligação obtidas experimentalmente, como uma soma de funções parametrizadas, como foi descrito por Böhm⁶⁷. O conceito na base deste tipo de funções é o pressuposto de que as energias livres de ligação podem ser aproximadas pela soma de termos, individuais e não correlacionados. Os coeficientes destes termos são obtidos por análise de regressão, utilizando energias livres de ligação e dados estruturais de ligandos conhecidos obtidos experimentalmente. Este facto faz com que as funções de *scoring* empíricas sejam altamente dependentes do grupo de dados utilizados para fazer a análise de regressão.

As funções empíricas podem incluir termos para contribuições não entálpicas, como o termo de rotor, que é uma aproximação aos efeitos entrópicos durante o reconhecimento molecular, e que é calculado pela soma ponderada do número de ligações do ligando que podem rodar⁶⁸. Outros termos incluídos podem ser de efeitos de solvatação e dessolvatação⁶⁹. No entanto, estes termos não descrevem totalmente os efeitos entrópicos e de solvatação na ligação proteína-ligando.

Funções de *scoring* baseadas em conhecimento estrutural. Funções deste tipo tentam reproduzir, preferencialmente, resultados estruturais obtidos experimentalmente, em vez de resultados de afinidade. Estas funções descrevem as interações proteína-ligando como simples modelos de potencial de pares atómicos de interação. Tal como as funções de *scoring* empíricas, as funções baseadas em conhecimento

estrutural tentam ter em conta efeitos no reconhecimento molecular que são difíceis de modelar explicitamente⁶⁶. A vantagem destas funções é a sua simplicidade computacional, o que facilita o teste de grandes bibliotecas de compostos virtuais. No entanto, o número de complexos proteína-ligando descritos estruturalmente, e cuja informação é usada para derivar este tipo de funções, é relativamente limitado.

1.4. Design de fármacos baseado em fragmentos químicos

O conceito de *design* de fármacos baseado em fragmentos químicos (FBDD) surgiu com os trabalhos de Jenks⁷⁰ e Ariens⁷¹, que propuseram que um fármaco poderia ser dividido em várias subestruturas moleculares, chamadas de epítomos ou fragmentos, que ligam a diferentes regiões do local de ligação, próximas entre si, e que a energia livre de ligação do fármaco seria resultado da soma das energias livres de cada epítopo. A primeira aplicação deste conceito no *design* de ligandos foi feita por Fesik e seus colaboradores⁷², numa abordagem usando SAR por NMR. Após cerca de 15 anos, surge o primeiro fármaco aprovado pela FDA, desenvolvido através de FBDD – o vemurafenib, vendido sob o nome comercial de Zelboraf – um agente quimioterapêutico contra melanoma expressando o gene *BRAF* mutado^{73, 74}. A aprovação deste fármaco constitui um marco crucial para a afirmação da potencialidade do FBDD.

Limitações da abordagem convencional e espaço químico de fragmentos

Na abordagem convencional ao *design* de fármacos, são testadas moléculas que, geralmente possuem características muito próximas de fármacos, algumas descritas pelas regras de Lipinski. Embora o teste deste tipo de moléculas com características de fármacos possa ser visto como vantajoso, na medida em que há maior probabilidade de estas

serem capazes de chegar aos alvos terapêuticos, o espaço químico de moléculas com massa molecular abaixo de 500 é cerca de 10^{60} , valor este que é muito difícil de amostrar, mesmo utilizando ensaios de HTS que conseguem testar milhares de compostos contra um determinado recetor⁷⁵⁻⁷⁷.

Outra limitação é o aumento da massa molecular (em cerca de 80) e lipofilicidade (aproximadamente 1 unidade logarítmica), resultante da adição de grupos químicos durante a otimização de moléculas com afinidade elevada para o alvo nos ensaios HTS⁷⁸. Estes aumentos resultam na perda de características de fármaco pelo ligando, reduzindo a probabilidade de este ser uma molécula biologicamente ativa⁷⁹.

Estas limitações, associadas à abordagem que envolve moléculas com características de fármacos, não existem numa abordagem FBDD. Numa campanha deste tipo são usadas bibliotecas de moléculas de pequenas dimensões, geralmente com massa molecular próxima de 250 e pouco funcionalizadas. Existem vários critérios para a construção de tais bibliotecas, sendo o conjunto de critérios mais comum a "Regra dos três"⁸⁰. Este conjunto de critérios foi proposto por Congreve e Carr, após uma análise de um conjunto variado de fragmentos com boas propriedades contra uma panóplia de alvos, e diz que um fragmento deve ter massa molecular não superior a 300, um número de doadores ou aceptadores de ligações de hidrogénio igual ou inferior a 3 e ClogP igual ou inferior a 3. Outros critérios que podem ser adicionados incluem o número de ligações que podem rodar não superior a 3 e superfície polar acessível não superior a 60.

A vantagem principal da abordagem baseada em fragmentos em relação às estratégias convencionais, é que o espaço químico de fragmentos com massa molecular não superior a 300 ($\sim 10^8$) é bastante menor, o que faz com seja mais fácil amostrar o espaço químico de fragmentos do que de moléculas com tamanho não superior a 500⁷⁵⁻⁷⁷.

Energia livre de ligação versus eficiência do ligando

No *design* de fármacos seguindo a abordagem clássica, a afinidade, que pode ser medida em termos de energia livre de ligação, é o fator mais importante na seleção dos melhores ligandos. No entanto, existe correlação entre a afinidade e a massa molecular, o que faz com que moléculas que tenham boas afinidades nos ensaios de HTS sejam demasiado grandes, violando mais do que uma regra de Lipinski.

A eficiência do ligando (LE), uma medida proposta por Hopkins e seus colaboradores⁸¹, permite contornar o enviesamento da afinidade para compostos maiores, ao contabilizar o contributo de cada átomo diferente de hidrogénio (cujo número é definido por HAC) para o total da energia livre de ligação. Esta medida é obtida da seguinte forma:

a partir da afinidade de ligação (K_d) dada pelos ensaios pode se obter a energia livre de ligação

$$\Delta G = -RT \ln K_d$$

a energia livre por átomo diferente de hidrogénio obtém-se a partir de

$$LE = \frac{\Delta G}{HAC}$$

Apesar de os ligandos selecionados através de um ensaio de HTS terem afinidades de ligação relativamente elevadas (K_d na ordem dos 100-1 μ M), em grande parte dos casos a LE é baixa⁷⁹. Uma explicação para este facto está descrita na Figura 2, que descreve, de uma forma simplificada, um ligando com afinidade relativamente elevada mas cuja ligação ao alvo é pouco específica, ou seja, pouco eficiente.

A LE é especialmente importante na abordagem FBDD. Como os fragmentos são pequenos e pouco funcionalizados, as afinidades destes são muito mais baixas do que as de ligandos testados em ensaios HTS⁸². A LE pode assim ser usada como guia para a adição de funcionalizações e otimização do fragmento, pois permite uma mais

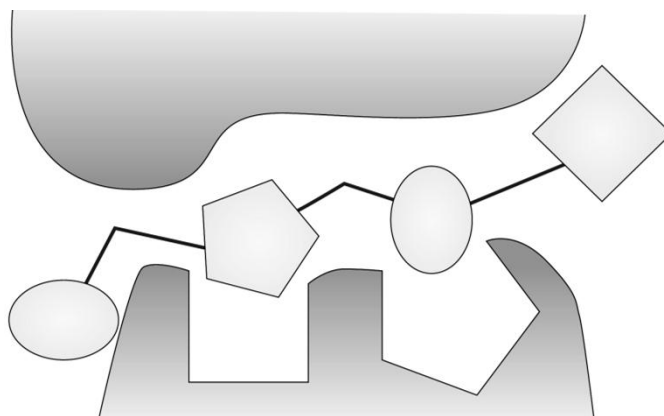


Figura 2. Representação esquemática de um ligando com características de fármaco com afinidade relativamente elevada para o alvo, mas com baixa LE. A afinidade deste tipo de moléculas deriva da superfície molecular em contacto com o recetor, sem que se formem interações de elevada eficiência com as regiões chave do local de ligação. Este tipo de moléculas, na ausência de informação estrutural (num estudo de SAR, por exemplo), torna-se muito difícil de otimizar. As Figuras 4 e 5 mostram o exemplo contrário, numa abordagem baseada em fragmentos químicos.

fácil racionalização da contribuição energética dos grupos químicos utilizados.

Embora a medida de LE seja um conceito bastante importante descoberta de novos fármacos, há uma clara tendência, transversal a todos os alvos, que indica que a LE diminui com o aumento do número de HAC. Esta tendência foi observada por Reynolds *et al.*⁸³, num estudo à biblioteca BindingDB⁸⁴, uma base de dados contendo informação sobre as constantes de afinidade de uma grande variedade de complexos ligando-alvo. Deste modo, torna-se difícil interpretar o valor de LE e compará-lo entre ligandos ou fragmentos de diferentes tamanhos, sendo este problema bastante evidente, devido ao crescimento dos fragmentos iniciais em candidatos com características de fármacos.

A medida que Reynolds e seu grupo propõem no mesmo estudo⁸³, é uma que incorpora o tamanho do ligando, em termos de número de HA, na medida de LE. Através do ajuste de uma função exponencial

aos maiores valores LE para cada valor de HAC, presentes na BindingDB, foi criado um fator de escala, ao qual chamou de LE_{Scale} , sendo este descrito pela seguinte equação:

$$LE_{Scale} = -0.064 + 0.873e^{(-0.026 \times HAC)}$$

A medida proposta chama-se *fit quality* (FQ) e é definida pela equação:

$$FQ = \frac{LE}{LE_{Scale}},$$

sendo esta medida faz com que os melhores ligandos tenham um valor de FQ próximo de 1, independentemente do tamanho.

Do fragmento ao candidato

Devido às baixas afinidades associadas aos fragmentos, a identificação destes no local de ligação do alvo é geralmente feita através de métodos biofísicos, como cristalografia de raios-X, NMR, SPR, entre outros, sendo que ensaios de HTS não são eficientes, nesta abordagem, devido aos seus baixos limites de detecção⁸⁵. Destas metodologias, a cristalografia de raios-X e, em alguns casos, o NMR são as mais aconselhadas, principalmente devido à informação sobre o modo de ligação ao alvo que pode ser obtida a partir destes métodos.

Após identificação dos fragmentos iniciais, é necessário “crescer” os fragmentos, de forma a aumentar a especificidade – que é reduzida nos fragmentos devido à baixa funcionalização destes – e a potência. Existem várias estratégias para transformar fragmentos iniciais em candidatos, sendo as mais importantes a evolução de fragmentos e a ligação de fragmentos.

Evolução de fragmentos. Este tipo de estratégia segue o conceito de adição de grupos para criar interações com outras regiões do local de ligação alvo, e consequentemente aumentar a potência, conceito este, bastante utilizado na abordagem convencional ao *design* de novos fármacos^{79, 85}. Esta estratégia é a mais popular e efetiva em FBDD. Quando existe informação estrutural do ligando, uma abordagem de

design com base na estrutura do alvo é utilizada⁸⁶. Caso contrário, são testados análogos dos fragmentos iniciais de forma a aplicar uma abordagem SAR⁸⁵.

Uma abordagem ao desenvolvimento de fragmentos, similar à evolução, é a otimização de fragmentos. Nesta abordagem, tal como na evolução de fragmentos, é feita a adição de grupos funcionais ao fragmento. No entanto, ao passo que na evolução o objetivo é aumentar a potência do fragmento, na otimização o objetivo é melhorar as propriedades, como a solubilidade ou permeabilidade. Um esquema do processo está representado na Figura 3

Ligação de fragmentos. Esta estratégia envolve a ligação de dois fragmentos que se ligam em regiões diferentes, mas próximas, do local de ligação alvo, usando um “ligador” molecular, ou *linker*. Este *linker* pode ser flexível ou rígido. Esta técnica tornou-se de certa forma sinónimo de *design* de fármacos com base em fragmentos, em muito devido ao facto que o primeiro trabalho de aplicação da abordagem FBDD, o trabalho “SAR by NMR” referido anteriormente⁷², utiliza a estratégia de ligação de fragmentos para desenvolver inibidores para a

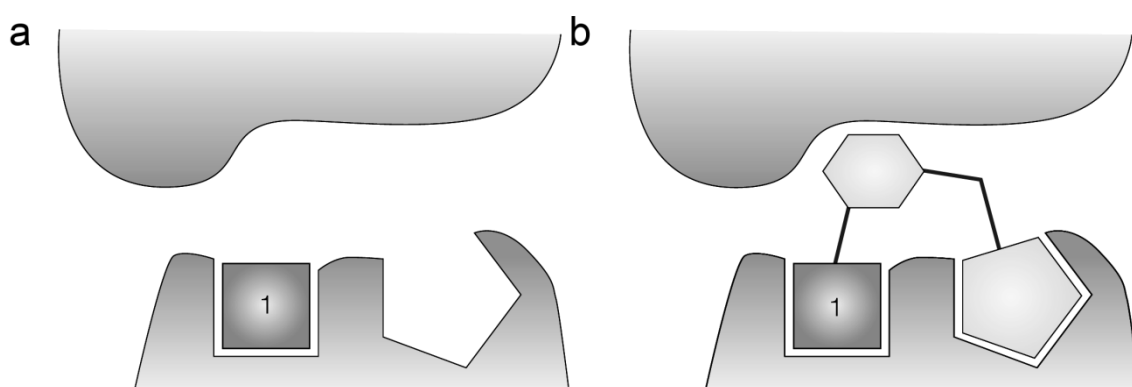


Figura 3. Representação esquemática da estratégia de evolução de fragmentos. a) o fragmento 1 liga-se uma região chave do local de ligação alvo; b) o fragmento 1 é evoluído através da adição de grupos substituintes de forma a estabelecer interações favoráveis com a superfície superior e a segunda região chave do local de ligação.

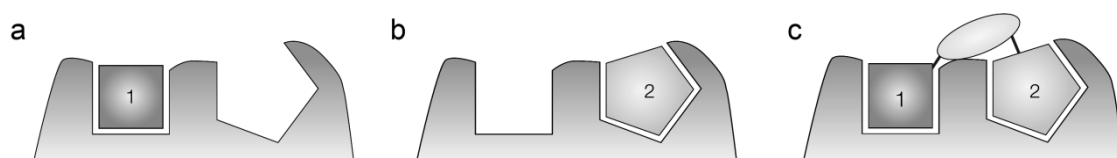


Figura 4. Representação esquemática da estratégia de ligação de fragmentos. a) o fragmento 1 liga-se na primeira região chave do local de ligação; b) o fragmento 2 liga-se numa segunda região chave, adjacente à região 1; c) os 2 fragmentos são ligados entre si através de um *linker*, sem que a conformação dos fragmentos 1 e 2 seja alterada.

FK506 *binding protein*. Na Figura 4 está representado um esquema que descreve a estratégia de ligação de fragmentos.

Quando a ligação de dois fragmentos é feita da forma ideal, ocorre um efeito de cooperatividade positiva, semelhante ao efeito referido no capítulo das interações intermoleculares, e referido como efeito de superaditividade. Isto significa que a soma das energias livres de ligação de cada fragmento separado será menos negativa que a energia livre de ligação do ligando resultante da ligação desses dois fragmentos entre si. A equação que descreve este efeito foi proposta por Jencks⁷⁰:

$$\Delta G_{AB} = \Delta G_A + \Delta G_B + \Delta G_s$$

em que ΔG_A e ΔG_B são as energias livres de fragmentos A e B, e ΔG_s é a energia livre da ligação, que pode ser vista como o termo de superaditividade. Este termo pode ser decomposto em vários termos energéticos que têm em conta as alterações entrópicas na rotação e translação, e as perdas na energia livre de ligação, associadas a alterações na conformação de ligação dos fragmentos A e B, que poderão ter ocorrido pela introdução do *linker*. Assumindo o caso ideal, em que não há alterações nas conformações de ligação de A e B, podemos assumir que o termo de superaditividade é definido pelas alterações na entropia de rotação e translação⁸⁷.

A constante de dissociação K_D^{AB} para dois fragmentos A e B ligados entre si pode ser descrita como

$$K_D^{AB} = K_D^A K_D^B E,$$

em que K_D^A e K_D^B são as constantes de dissociação de A e B, respetivamente, e E é um coeficiente de ligação, também descrito na literatura como fator teórico do *linker* de fragmentos, f_L ⁸⁸. Este coeficiente constitui uma medida útil da eficiência da ligação de dois fragmentos, e está relacionado com a equação de ligação de fragmentos, proposta por Jencks da seguinte forma:

$$\Delta G_s = RT \ln E$$

Coeficientes de ligação abaixo de um significam que existe um efeito de cooperatividade positiva, enquanto um coeficiente de ligação acima de 1 significa que existe um efeito de cooperatividade negativa, isto é, que a soma das energias livres dos fragmentos A e B é maior que a energia livre do composto resultante da ligação dos dois.

A principal limitação desta estratégia está na escolha do *linker*. Um *linker* rígido poderá induzir alterações na ligação dos fragmentos ligados, ao passo que um *linker* flexível levará provavelmente a uma penalização do ponto de vista entrópico.

Metodologias computacionais em FBDD

Como já foi referido, a natureza intrínseca dos fragmentos químicos usados em FBDD faz com que os testes contra os alvos terapêuticos tenham de ser feitos através de metodologias biofísicas, como a cristalografia de raios-X ou o NMR. Apesar de estas metodologias permitirem ao químico medicinal conhecer a estrutura dos complexos fragmento-alvo, os custos, morosidade e débito destas são fatores desfavoráveis para uma abordagem unicamente experimental à estratégia de FBDD. Outra desvantagem está no facto de nem todos os alvos serem cristalizáveis, ou possíveis de estudar usando NMR.

Uma forma de contornar estas desvantagens da aplicação experimental do FBDD, é recorrer a métodos computacionais. O *docking*, bastante utilizado em campanhas de *design* de fármacos com base na estrutura do alvo⁶⁶, também já foi aplicado com sucesso na identificação de fragmentos com potencial para alguns alvos

terapêuticos, como a trombina⁸⁹, fator Xa⁹⁰ e cinase dependente de ciclinas 4 (CDK4)⁹¹. No entanto a construção de bibliotecas de fragmentos e a aplicação do *docking* em FBDD são complexas e ainda há progressos a fazer nesta área.

Uma das dificuldades está no facto de algumas funções terem sido parametrizadas de acordo com bibliotecas de ligandos com características de fármacos. Alguns estudos recentes, no entanto, apontam para que não existam diferenças significativas no desempenho das metodologias de *docking*, quer para moléculas com características de fármacos, quer para fragmentos químicos^{77, 92}. Chen e Soichet⁷⁷ publicaram um estudo em que encontraram 10 fragmentos com elevada afinidade (no contexto dos fragmentos) para a β -lactamase CTX-M, a partir de 69 fragmentos testados por *docking*. As estruturas de cristalografia de raios-X subsequentes, resolvidas por estes autores, mostraram que as conformações de ligação previstas correspondiam ao obtido experimentalmente. Estudos como estes são fortes evidências de que a aplicação de metodologias computacionais em FBDD é uma mais-valia. No entanto não existe nenhum estudo sistemático, isto é, sobre um conjunto variado de alvos, que confirme estas evidências, o que é importante dada a heterogeneidade de locais de ligação alvos com potencial terapêutico.

Este estudo teórico vem preencher essa ausência de um estudo sistemático. Neste trabalho, através do *docking* de uma biblioteca de 5930 fragmentos – criada com base em grupos químicos frequentes em fármacos comercializados – contra um conjunto variado de 35 alvos, respondemos a três questões chave para o FBDD: 1) até que ponto é, a biblioteca de fragmentos criada, capaz de amostrar as regiões chave do local ativo dos alvos? 2) Será, a metodologia de *docking* usada, capaz de reproduzir de forma precisa a conformação de ligação dos epítomos dos ligandos experimentais? 3) Qual o grau de promiscuidade de cada fragmento contra cada elemento do conjunto de alvos?

Métodos

2. Métodos

2.1. Construção da biblioteca de fragmentos

A biblioteca de fragmentos foi construída usando o mesmo conjunto de núcleos, todos eles aromáticos com exceção do ciclohexano e piperidina, e cadeias laterais propostos por Hartshorn⁹³ e seus colaboradores, e representados na Figura 5. Estas estruturas foram selecionadas por serem frequentes em fármacos conhecidos, como demonstrado por estudos como os de Bemis e Murcko^{94, 95}. Cada fragmento resulta da combinação de um núcleo com uma cadeia lateral. Todos os carbonos foram substituídos com uma cadeia lateral, e, no caso de centros quirais, ambas as posições foram exploradas. Átomos de nitrogénio heterocíclicos apenas foram substituídos com grupos metilo, acetilo ou metil sulfona. Foram adicionados todos os estados protonação de grupos ionizáveis no intervalo de pH entre 5 e 9. Os valores de pKa foram calculados utilizando a ferramenta Marvin Calculator Plugins (versão 5.8.0) disponibilizada pela ChemAxon⁹⁶. A biblioteca virtual de fragmentos resultantes contém um total de 5930 moléculas, maior do que os 4513 fragmentos da biblioteca desenvolvida por Hartshorn⁹³. O maior tamanho da biblioteca deve-se à inclusão dos diferentes estados de protonação dos fragmentos com grupos ionizáveis. A parametrização dos fragmentos foi feita através da otimização de geometria através de uma abordagem de mecânica quântica ao nível de Hartree-Fock com o basis set 6-31G(d) utilizando o programa GAMESS⁹⁷. As geometrias otimizadas de cada fragmento e o seu potencial eletrostático foram utilizados para calcular cargas atômicas parciais através do método de ajuste RESP⁹⁸.

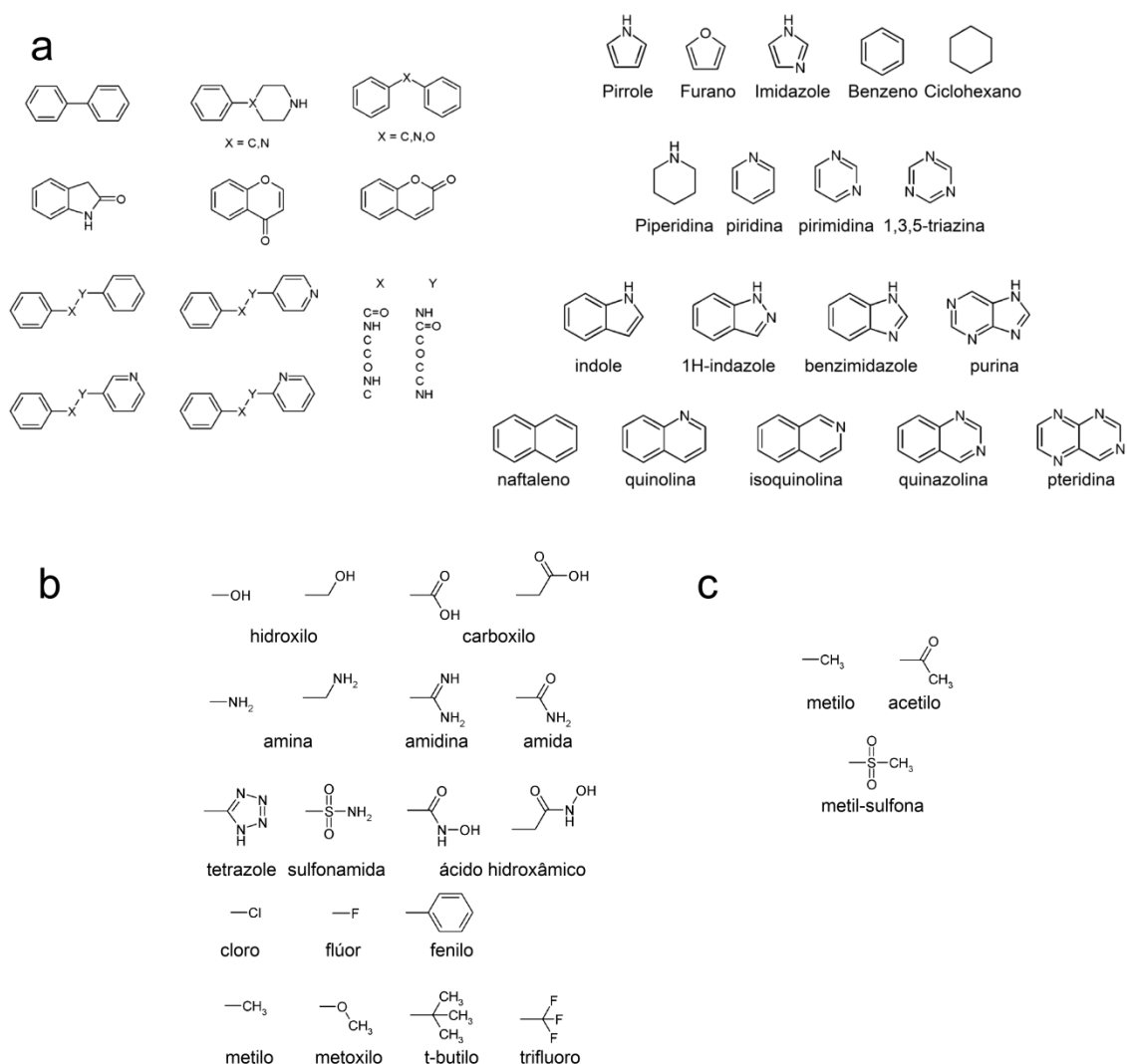


Figura 5. Núcleos e cadeias laterais utilizadas para a construção da biblioteca de fragmentos. Cada fragmento resulta da combinação de uma núcleo e uma cadeia lateral. Estas estruturas são as mesmas utilizadas por Hartshorn e colaboradores, e são estruturas frequentemente encontradas em fármacos. a) seleção de núcleos utilizados na construção da biblioteca de fragmentos; b) cadeias laterais para a substituição de átomos de carbono; c) cadeias laterais para a substituição de átomos de nitrogénio.

2.2. Seleção de alvos

Foram selecionados 32 alvos de um total de 85 que fazem parte do Astex Diverse Set⁹⁹. O critério de seleção para estes 32 alvos foi a semelhança dos epítomos dos ligandos com alguns dos fragmentos presentes na biblioteca de fragmentos utilizada neste trabalho. Esta

Tabela 2. Alvos selecionados para as experiências de *docking* da coleção Astex Diverse Set⁹⁹. Estão também representados os ligandos respectivos a cada alvo e os epítomos considerados destacados em caixas azuis para epítomos A, vermelhas para epítomos B e verdes para epítomos C.

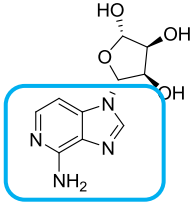
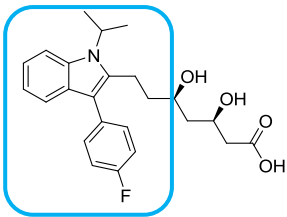
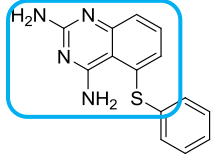
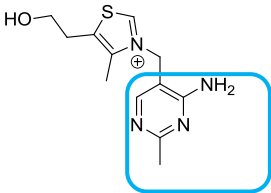
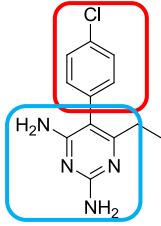
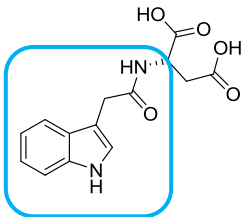
Código PDB	Alvo	Estrutura do Ligando	Medida de Potência	Potência (μM)
1HP0	purine specific nucleoside hydrolase		K _i	0,20000
1HWI	HMG-coA reductase		IC ₅₀	0,02800
1IA1	dihydrofolate reductase		IC ₅₀	0,03400
1IG3	thiamin pyrophosphokinase			
1J3J	dihydrofolate reductase		K _i	0,00980
1K3U	tryptofan synthase			

Tabela 2. (continuação)

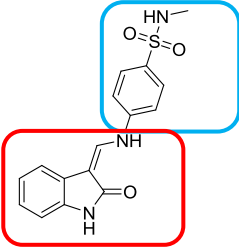
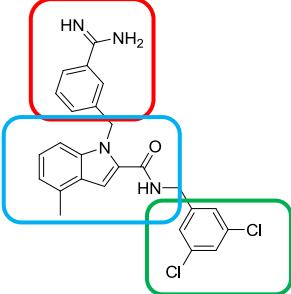
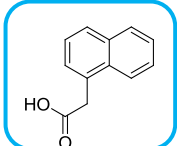
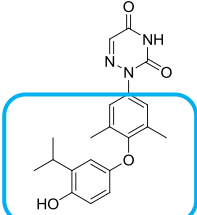
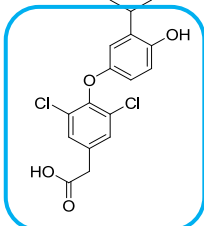
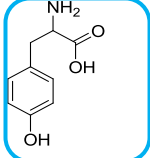
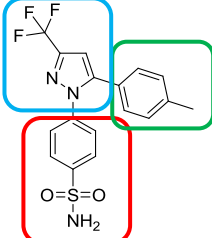
1KE5	cyclin-dependent kinase 2		IC ₅₀	0,56000
1LPZ	factor Xa		K _i	0,02500
1LRH	auxin-binding protein 1			
1N46	thyroid hormone receptor β 1		K _i	0,00003
1NAV	thyroid hormone receptor α1		IC ₅₀	0,02500
1OF6	DAHP synthase		K _i	0,90000
1OQ5	carbonic anhydrase II		IC ₅₀	0,00210

Tabela 2. (continuação)

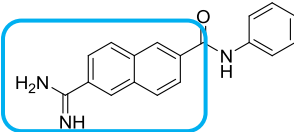
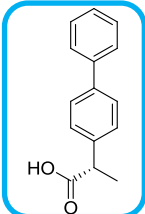
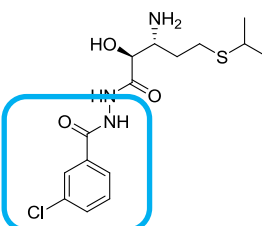
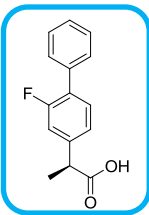
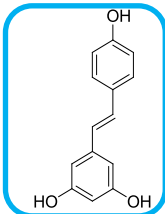
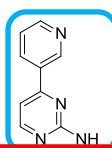
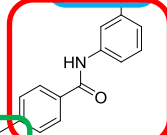
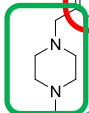
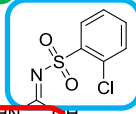
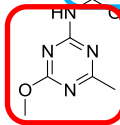
1OWE	urokinase		K _i	0,63100
1Q4G	prostaglandin H2 synthase 1		K _i	0,13000
1R58	methionine aminopeptidas e 2		IC ₅₀	0,11000
1R9O	cytochrome P450 2C9		K _i	9,60000
1SG0	quinone reductase		K _i	0,03400
1T46	C-kit tyrosine kinase	  	IC ₅₀	0,41300
1T9B	acetohydroxy acid synthase	 	K _i	0,12700

Tabela 2. (continuação)

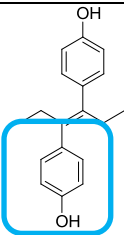
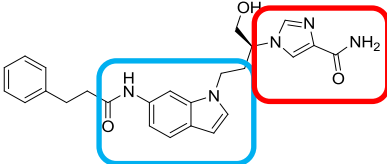
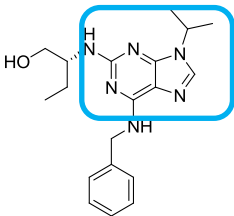
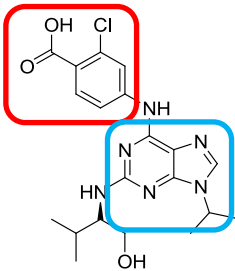
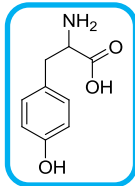
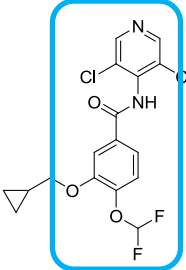
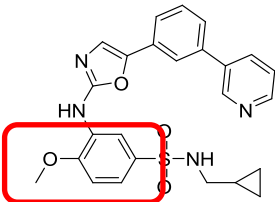
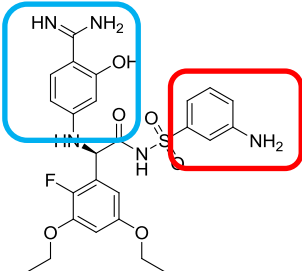
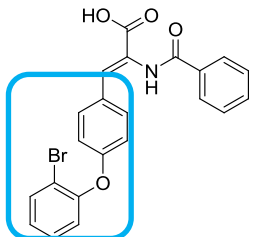
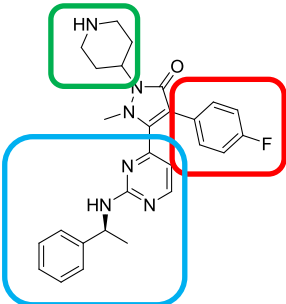
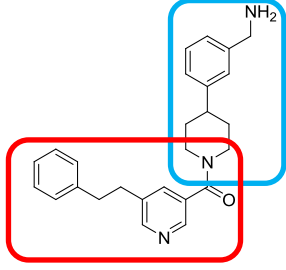
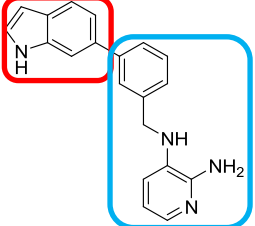
1TZ8	transthyretin			
1UML	adenosine deaminase		K _i	0,03000
1UNL	cyclin-dependent kinase 5		IC ₅₀	0,20000
1V0P	protein kinase 5		IC ₅₀	0,13000
1X8X	tyrosyl-tRNA synthase			
1XOQ	phosphodiesterase 4D		IC ₅₀	0,03800
1Y6B	vascular endothelial growth factor receptor 2			

Tabela 2. (continuação)

1YGC	factor VIIa		K _i	0,00035
1YVF	NS5B polymerase		IC ₅₀	0,10000
1YWR	p38 kinase		IC ₅₀	0,03200
2BM2	βII tryptase		K _i	0,01500
2OHT	BACE-1		IC ₅₀	9,10000

seleção de alvos, e os ligandos correspondentes, com os epítodos assinalados estão representados na tabela 2. Neste estudo, definimos epítodo como uma subestrutura de um dado ligando que seja similar a, pelo menos, um dos fragmentos da nossa biblioteca, sendo que um ligando pode ter até três epítodos. Os epítodos são considerados

semelhantes aos fragmentos se possuírem um núcleo cíclico e cadeias laterais semelhantes às utilizadas na construção da biblioteca de fragmentos. Em dois casos, os alvos 1LRH e 1TZ8, os ligandos são completamente iguais a fragmentos da nossa biblioteca.

As regiões chave dos locais de ligação alvo são definidas pelos locais onde se ligam os epítomos considerados, e, tal como estes últimos, o seu número varia entre um e três.

2.3. Docking

As experiências de *docking* foram feitas utilizando o programa Autodock 4¹⁰⁰⁻¹⁰². Este programa possui um algoritmo de procura Lamarckiano, consistindo num algoritmo genético, como o explicado anteriormente, juntamente com um método de minimização de energia. A função de *scoring* é baseada num campo de forças semiempírico, que incorpora a energia interna do ligando na energia livre de ligação prevista, e um método de cálculo da dessolvatação baseado nas cargas atômicas parciais, e que foi calibrada para um conjunto de 288 complexos alvo-ligando conhecidos.

Os fragmentos da biblioteca foram considerados totalmente flexíveis, isto é, foi permitida a rotação de todos os diedros próprios. Os tipos de átomos dos alvos proteicos foram os definidos no campo de forças do Autodock 4, e as cargas atômicas parciais para estes foram calculadas utilizando o método de Gasteiger-Marsili^{103, 104}.

A grelha, utilizada nos cálculos de energia sonda-alvo, foi colocada com o seu centro no local de ligação alvo, identificado pela presença do ligando experimental, sendo que o seu tamanho foi definido manualmente de forma a incluir todas as regiões do local de ligação de cada alvo. As maiores grelhas apresentavam dimensões 25x28x29 pontos, espaçados entre si em 0,375 Å. Para cada fragmento foram efetuadas 10 corridas do algoritmo genético, com 150 indivíduos em cada população. Para cada corrida, o número máximo de gerações foi definido como 27x10³ e o número máximo de avaliações foi 5x10⁵. As

soluções de *docking* resultantes foram submetidas a uma análise de *clustering* com o AUTODOCK, com o parâmetro de corte *rtol* igual a 1 Å. Do *cluster* de soluções para o *docking* de cada fragmento contra cada um dos alvos, apenas os complexos alvo-fragmento com a menor energia foram selecionados.

2.4. Análises

Os fragmentos foram classificados segundo duas métricas referidas anteriormente: a LE (eficiência do ligando) e o FQ (*fit quality*). Estas medidas foram calculadas sobre a energia livre de ligação calculada pelo Autodock 4. Assim, os fragmentos ditos “melhores”, serão os que estarão no topo da classificação, para cada uma das métricas.

Devido ao facto de os epítomos e os fragmentos correspondentes não serem exatamente iguais – alguns epítomos apresentam mais do que uma funcionalização – foi necessário utilizar uma medida de RMSD sem correspondência direta entre os átomos. A medida de RMSD proposta neste trabalho, para duas estruturas, a e b, similares entre si, define-se da seguinte forma:

$$RMSD_{ab} = \max(RMSD'_{ab}, RMSD'_{ba})$$

em que

$$RMSD'_{ab} = \sqrt{\frac{1}{N} \sum_i \min_j r_{ij}^2}$$

O somatório é sobre todos os N átomos diferentes de hidrogénio na estrutura a e o mínimo é sobre todos os átomos na estrutura b. r^2 é o quadrado da distância entre os átomos *i* e *j*.

Todas as figuras foram criadas utilizando o programa Pymol¹⁰⁵.

Resultados e discussão

3. Resultados e discussão

Neste trabalho propomos responder, de uma forma sistemática, a quatro questões fundamentais associadas à compreensão, validação e aplicação de metodologias de *docking* molecular e bibliotecas de fragmentos em FBDD. Estas questões têm como objetivo aumentar a nossa compreensão acerca da capacidade das bibliotecas de fragmentos amostrarem o local de ligação alvo, da precisão das funções de procura e de *scoring* das soluções de *docking*, perceber a seletividade ou promiscuidade destes num conjunto variado de alvos, e inferir sobre a aplicabilidade de estratégias de ligação de fragmentos num contexto de *docking* molecular.

Para responder a estas questões desenvolvemos uma biblioteca de 5930 fragmentos, de acordo com Hartshorn *et al*⁹³, baseada em grupos funcionais e núcleos cíclicos frequentes em fármacos. Esta opção permite uma amostragem racional do espaço de fragmentos, ao utilizar grupos cujas propriedades são exploradas em moléculas com capacidade terapêutica, e evitando grupos com possível toxicidade inerente.

Para testar esta biblioteca utilizamos um conjunto diverso de 32 estruturas cristalográficas, de elevada resolução, de proteínas complexadas com ligandos relevantes do ponto de vista terapêutico, e com subestruturas similares a fragmentos da nossa biblioteca. Estas 32 estruturas são uma seleção de uma biblioteca criada pela Astex, e que foi desenvolvido para a validação de metodologias de *docking*⁹⁹. A seleção das 32 estruturas foi feita de acordo com a semelhança entre epítomos dos ligandos experimentais e elementos da nossa biblioteca de fragmentos.

3.1. Amostragem de regiões chave do local de ligação

Em qualquer campanha de FBDD, é necessário construir uma biblioteca de fragmentos que será posteriormente testada contra um,

ou uma série de alvos, experimentalmente ou computacionalmente. As bibliotecas utilizadas em *docking* de fragmentos são geralmente criadas com o objetivo de amostrar, da melhor forma possível, o espaço químico de fragmentos moleculares. No entanto, surge a questão se, de facto, a biblioteca criada é capaz de amostrar as regiões chave do local de ligação, isto é, se é possível, através do *docking* da biblioteca de fragmentos, obter soluções nas regiões chave do local de ligação de uma série de alvos. Neste trabalho, as regiões chave do local de ligação são definidas pelos epítomos dos ligandos experimentais, presentes nas estruturas cristalográficas utilizadas, como está representado na Figura 6.

Para medir a distância entre as soluções de *docking* e os epítomos dos ligandos experimentais foi calculado o RMSD. As soluções de *docking* utilizadas foram as melhores conformações do *cluster* de soluções de menor energia livre de ligação, uma para cada fragmento. Para todas as experiências de *docking* foi observada uma média de 5.3,

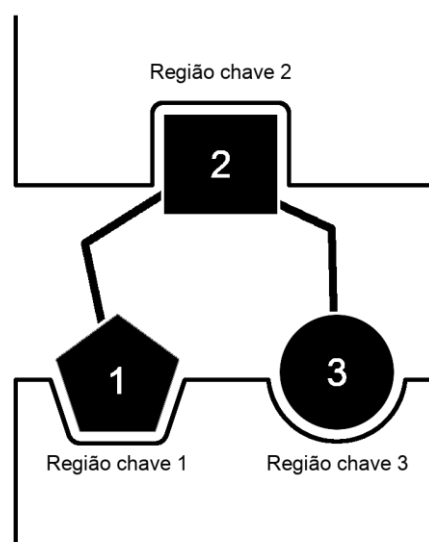


Figura 6. Representação esquemática das regiões chave do local de ligação consideradas. Cada região chave é definida pelo epítomo do ligando experimental que a ocupa. O cálculo de RMSD das soluções de *docking* de fragmentos foi feito contra cada um dos epítomos, separadamente. Valores de RMSD baixos significarão que as respectivas soluções estarão próximas da região chave ocupada pelo epítomo correspondente.

em 10, conformações no *cluster* de soluções com a menor energia livre de ligação. Para ligandos definidos por mais do que um epítipo, o RMSD dos fragmentos foi calculado para cada um, separadamente. A distribuição de fragmentos por intervalos de RMSD em 5 alvos selecionados está representada na Figura 7. Com esta distribuição de RMSD é possível obter uma primeira impressão sobre a localização dos fragmentos em relação a cada um dos epítipos do ligando experimental. A Figura 7a mostra que a grande maioria das soluções do *docking* de fragmentos contra o alvo 1TZ8 está num intervalo de baixo RMSD, isto é, está muito próxima do único epítipo considerado do ligando experimental, enquanto no caso do alvo 1HWI, que também possui apenas um epítipo, os fragmentos não possuem uma

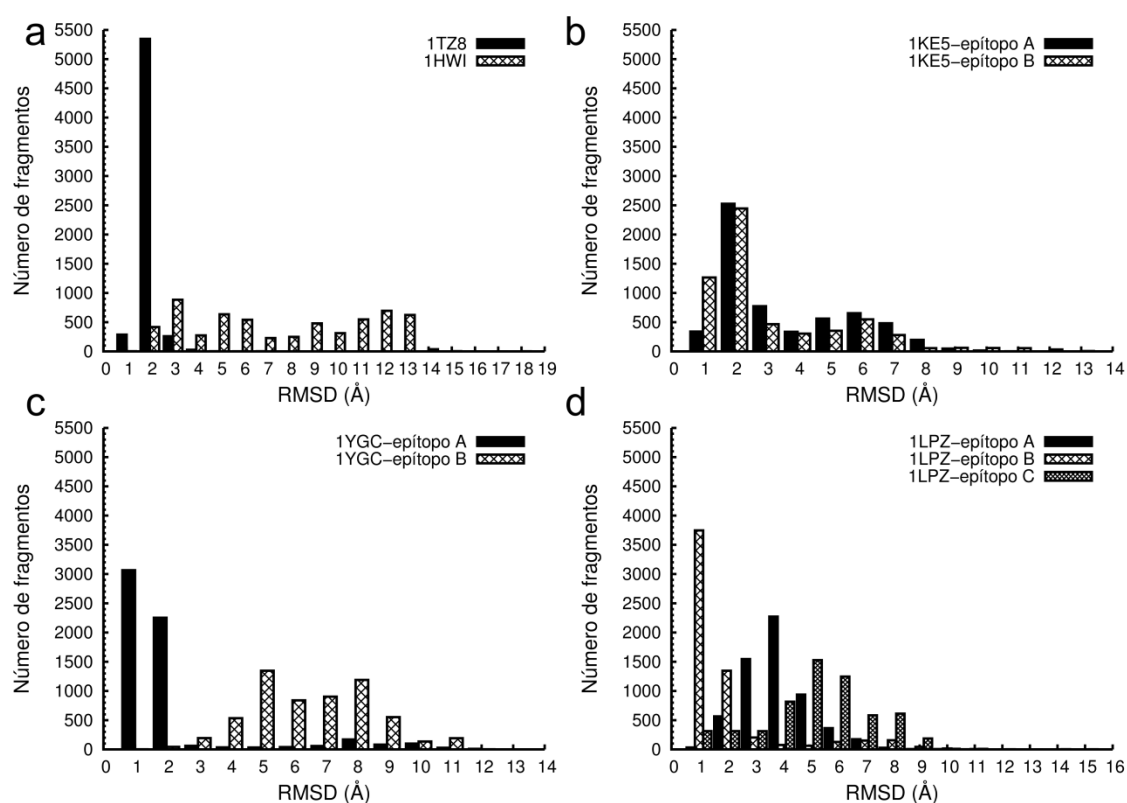


Figura 7. Número de soluções de *docking* de fragmentos a diferentes intervalos de RMSD em relação a um dado epítipo de ligando experimental, para 5 alvos selecionados. Ligandos com mais do que um epítipo estão legendados. a) alvos 1TZ8 e 1HWI; b) epítipos A e B do alvo 1KE5; c) epítipos A e B do alvo 1YGC e d) epítipos A, B e C do alvo 1LPZ.

preferência evidente para a região do local de ligação onde esse epítopo se encontra. A localização das conformações de *docking* de todos os fragmentos para os alvos 1TZ8 e 1HWI estão ilustradas na Figura 8a e Figura 8b, respetivamente. Cada ponto amarelo nas imagens representa o centro geométrico de cada fragmento. No caso do alvo 1TZ8, é claramente observável que as soluções de *docking* estão concentradas perto da referência experimental (Figura 7a), enquanto, no caso do alvo 1HWI, as soluções de *docking* estão dispersas pelo local de ligação (Figura 8b). A Figura 7b mostra o caso do alvo 1KE5, no qual as regiões de ambos os epítopos selecionados do ligando experimental possuem um elevado número de soluções de *docking* de fragmentos. Esta observação é evidente na Figura 8c, em que as soluções estão próximas de ambos os epítopos. Um caso oposto é o caso do alvo 1YGC (Figura 7c), em que o *docking* não colocou um número significativo de soluções nas proximidades de um dos epítopos, o epítopo B, e uma preferência evidente foi observada para a região do epítopo A. A elevada densidade de soluções na região do epítopo A é claramente observável na Figura 8d. A mesma figura mostra também a baixa densidade de soluções de *docking* na região do epítopo B. Ilustramos também o caso do alvo 1LPZ. O ligando associado a este alvo foi definido por três epítopos, cada um numa região chave do local ativo diferente. A maioria das conformações de *docking* dos fragmentos estão localizadas, preferencialmente, na mesma região que o epítopo B, como se pode observar Figura 7d. Este gráfico aparenta também alguma preferência para as regiões dos epítopos A e C. No entanto, a Figura 8e apresenta uma visão clara da localização das conformações de *docking* dos fragmentos ao longo das três regiões chave do local de ligação do alvo, verificando-se uma maior preferência pela região definida pelo epítopo B, seguida da região C e, finalmente, a região A.

Os nossos resultados mostram que, com a nossa biblioteca, é possível obter soluções de *docking* que vão de acordo com a

localização dos epítomos dos ligandos experimentais no local de ligação. De um ponto de vista de amostragem do local de ligação, é essencial que uma biblioteca de fragmentos seja capaz de amostrar, pelo menos, uma região chave do local de ligação alvo, a partir do qual um, ou mais fragmentos pode, ser “crescidos” a moléculas de elevada afinidade.

A análise dos 32 alvos (anexo I) sugere que, pelo menos, uma região

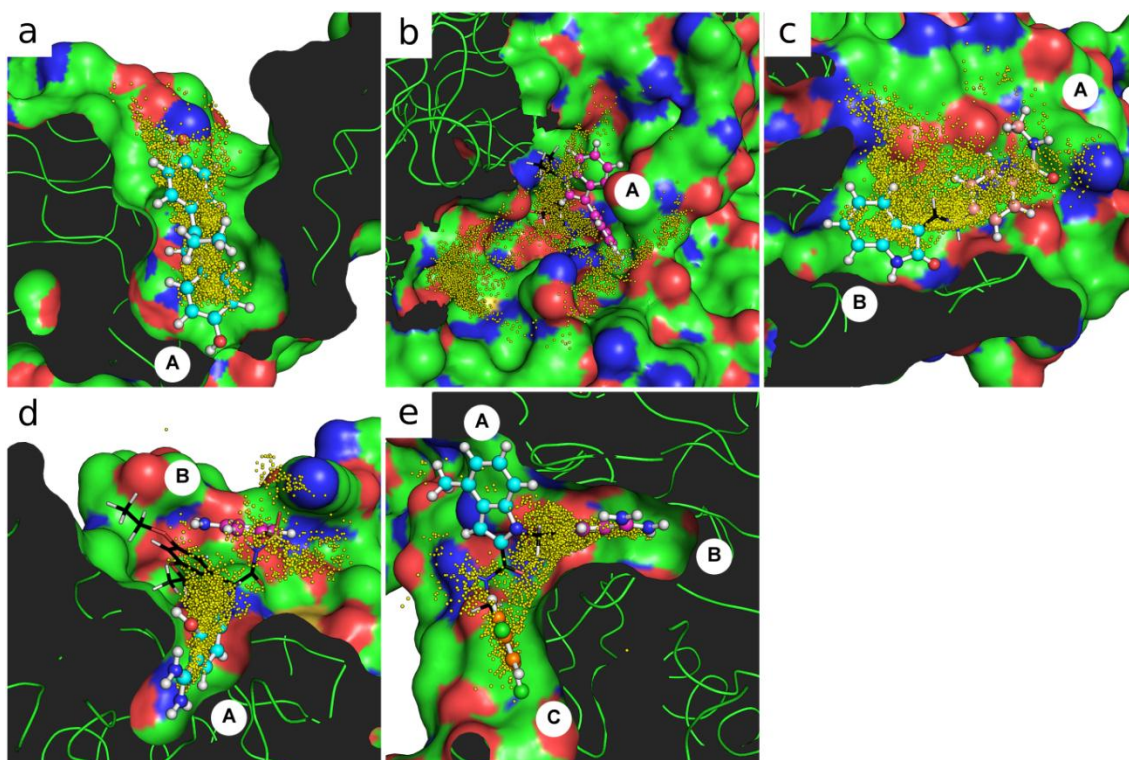


Figura 8. Localização do centro geométrico, representado em pequenas esferas amarelas, de todas as soluções de *docking* de fragmentos no local de ligação de: a) 1TZ8; b) 1HWI; c) 1KE5; d) 1YGC e e) 1LPZ. Os ligandos experimentais estão representados *sticks* com coloração CPK e cada epítipo está representado na forma de *ball-and-stick* com um esquema de cores diferente (no qual as cores dos carbonos variam entre ciano, magenta, laranja e rosa). A proteína está representada em *ribbons* de cor verde e uma superfície molecular com o seguinte esquema de cores: nitrogénio – azul, carbono – verde, oxigénio – vermelho. Os epítomos e as regiões chave do local de ligação correspondentes estão assinalados como A, B e C. Foi aplicado um corte transversal da proteína para facilitar a visualização do local de ligação.

chave é bem amostrada para todos os alvos com uma, duas ou três regiões chave do local de ligação. O exemplo que parece ter resultados menos satisfatórios é o alvo 1HWI (Figura 7a). Neste caso, a maior parte das soluções de *docking* estão espalhadas pelo local de ligação, sendo que a população mais próxima do epítipo do ligando experimental está localizada a 2-3 Å do seu centro geométrico. É de notar que o local de ligação deste alvo é exposto ao solvente e, e as bolsas hidrofóbicas que o compõem são pouco profundas. A Figura 7a ilustra também o caso do alvo 1TZ8, no qual a maioria das soluções de *docking* está localizada no local de ligação com forma de funil (Figura 8a). Observamos também que as bibliotecas de fragmentos conseguem amostrar duas regiões chave distintas num local de ligação, como acontece com o alvo 1KE5 (Figura 7b e Figura 8c). Neste caso, a biblioteca de fragmentos permitiu obter dois conjuntos de soluções de *docking* distintos, nas mesmas regiões ocupadas pelos epítipos experimentais. Este exemplo mostra claramente que bibliotecas de fragmentos quimicamente diversas podem ser desenvolvidas de forma a amostrar as propriedades químicas específicas de cada região chave de um determinado local de ligação, como por exemplo, se uma determinada região possui maior preferência para moléculas com substituintes apolares, ou substituintes com dadores/aceitadores de ligações de hidrogénio.

O alvo 1YGC é um exemplo de uma boa amostragem de uma região chave em detrimento de uma segunda região (Figura 7c e Figura 8d). Um caso semelhante, mas mais complexo, é o caso do alvo 1LPZ, descrito na Figura 7d e Figura 8e. Neste último observa-se claramente que existe uma hierarquia no nível de amostragem ao longo das três regiões chave do local de ligação, isto é, uma região apresenta um maior número de soluções de *docking* de fragmentos e que vai diminuindo sequencialmente para as restantes duas regiões. Estes dois casos (1YGC e 1LPZ) sugerem que os resultados de uma campanha de *docking* de fragmentos podem ser utilizados para, não

só dar prioridade a um determinado grupo de fragmentos para ser desenvolvidos, mas também dar prioridade às regiões chave que devem ser investigadas numa campanha de FBDD.

Estes exemplos, descritos na Figura 7 e na Figura 8, mostram que as bibliotecas podem ser enviesadas para uma determinada região chave no local de ligação, como acontece no caso do alvo 1YGC, o que leva a uma sobreamostragem de uma região específica, ou podem amostrar de forma equilibrada mais do que uma região chave do local de ligação do alvo, como acontece no caso do alvo 1KE5.

3.2. RMSD dos fragmentos contra epítomos de ligandos

A segunda questão deste trabalho foi estimar o desempenho da metodologia de *docking* na previsão da conformação dos fragmentos, em comparação com os epítomos dos ligandos experimentais. Esta análise é diferente da análise convencional de desempenho das metodologias de *docking*, na qual é feito o *re-docking*, isto é, a comparação da conformação experimental do ligando no alvo e a conformação desse mesmo ligando obtida com o *docking*. Neste trabalho a metodologia de *docking* é desafiada a colocar fragmentos da nossa biblioteca na mesma conformação que os epítomos de ligandos experimentais de elevadas afinidades. Os valores de RMSD dos epítomos experimentais em relação aos fragmentos equivalentes estão representados no gráfico da Figura 9. Ligandos com mais do que um epítopo estão identificados como a, b e c em expoente após o código do PDB, no eixo das abcissas. A distribuição destes valores de RMSD (gráfico do canto superior direito da Figura 9) mostra a presença de 23 fragmentos com um RMSD inferior a 1.85 Å. Tendo em conta que temos uma seleção de 45 epítomos, este valor corresponde a uma precisão de *docking* de 51.1%. Os restantes fragmentos com menor desempenho possuem um RMSD que vai até 11 Å do epítopo correspondente. Estes dados indicam que, em 51,1% dos casos, a metodologia de *docking* prevê a conformação de ligação do fragmento corretamente (com um

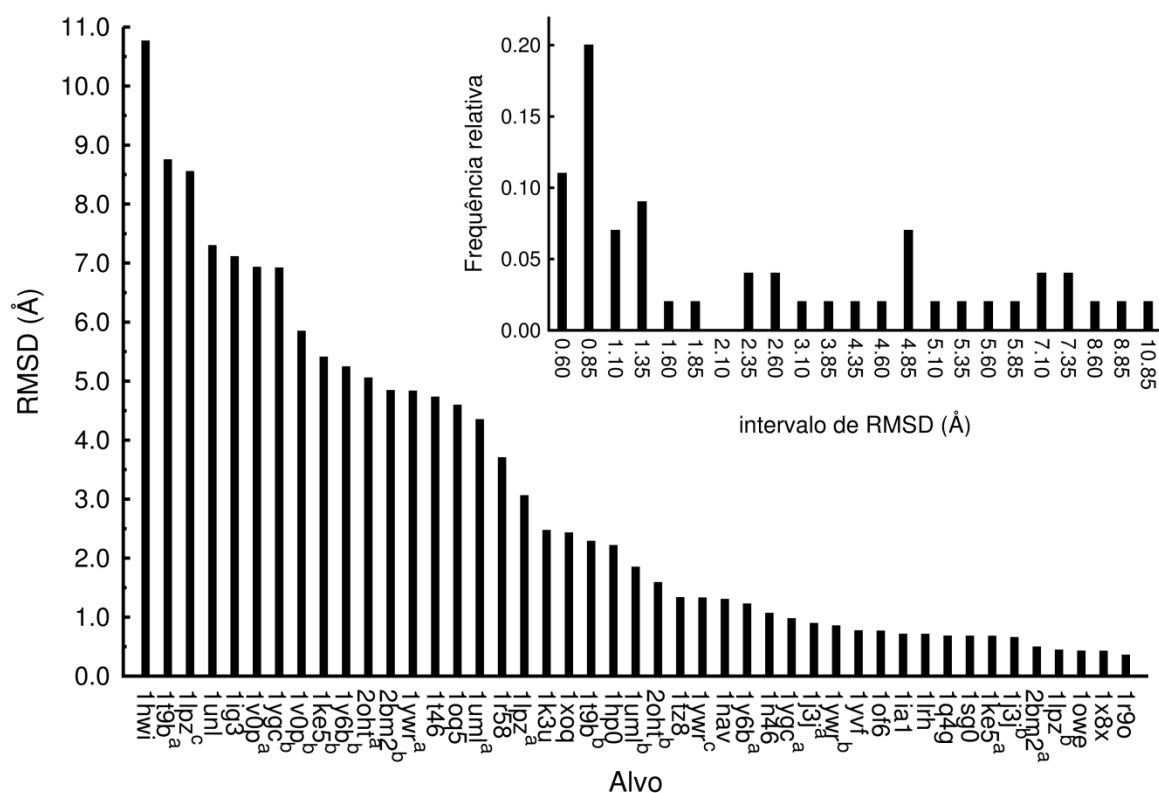


Figura 9. RMSD entre o epítipo experimental e o fragmento da biblioteca equivalente. Os alvos cujos ligandos com mais do que um epítipo estão assinalados com a, b ou c sobrescrito. No canto superior direito está representado o gráfico da distribuição de valores de RMSD.

RMSD não superior a 1.85 Å) contra, pelo menos, uma região chave do local de ligação. Esta percentagem diminui para 46.7 se assumirmos um limite de 1.50 Å, como sugerido por Verdonk *et al*⁹². De qualquer das formas, estas percentagens estão dentro do intervalo de precisão de *docking* obtida utilizando outros vários programas avaliados por Verdonk e seus colaboradores⁹². É de notar, no entanto, a concordância entre as conformações de fragmentos previstas pela metodologia de *docking* utilizada neste trabalho e a conformação dos epítipos de ligandos experimentais é um ponto forte a favor da utilização de fragmentos, testados através de *docking*, para futuro desenvolvimento.

A Figura 10 mostra alguns epítipos dos ligandos experimentais, e as conformações, obtidas através de *docking*, dos respetivos fragmentos

equivalentes. A Figura 10a-c apresenta as conformações de *docking* de fragmentos com o valor de RMSD em relação ao epítipo correspondente mais baixo, correspondendo aos alvos 1R9O, 1X8X e 1OWE. O fragmento ligado ao alvo 1R9O é o que apresenta a conformação mais próxima da do epítipo experimental, e estabelece as mesmas interações eletrostáticas e ligações de hidrogénio com a ASN-174 e ARG-78 (Figura 10a). O fragmento aromático ligado ao alvo 1X8X está em total sobreposição com o epítipo do ligando experimental e estabelece as mesmas ligações de hidrogénio com os resíduos de ASP-182 e TYR-37 (Figura 10b). O fragmento ligado ao alvo

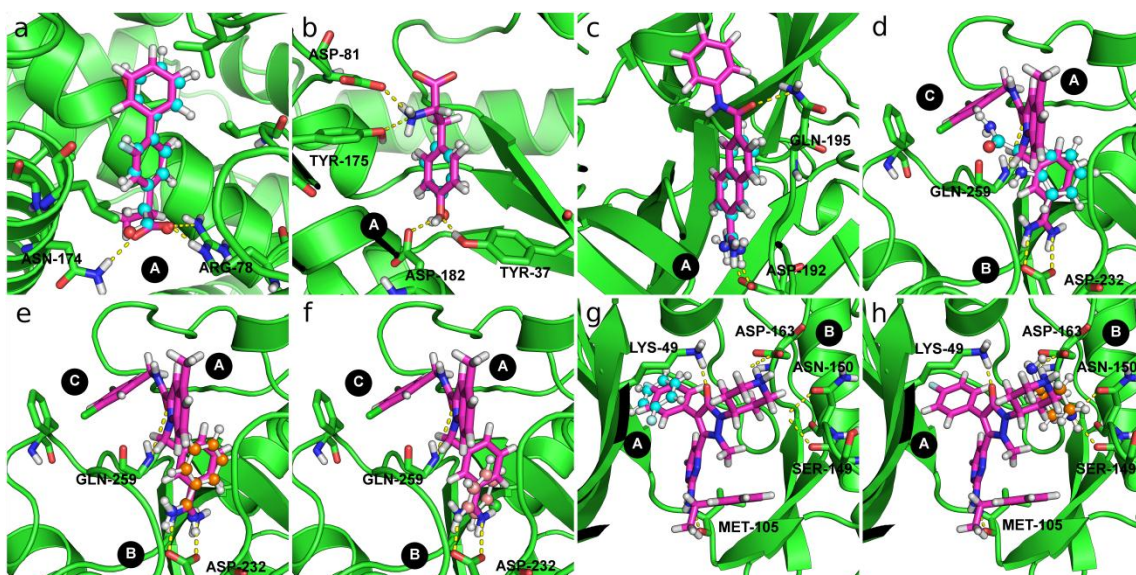


Figura 10. Conformações de ligação de ligandos experimentais dos alvos 1R9O, 1X8X, 1OWE, 1LPZ-A, 1LPZ-B, 1LPZ-C, 1J3J-A, 1J3J-B e os fragmentos da biblioteca equivalentes a cada um destes. Os ligandos experimentais estão representados com *sticks* com o seguinte esquema de cores: nitrogénio – azul, carbono – magenta, oxigénio – vermelho e hidrogénio – branco. Os fragmentos estão representados na forma *ball-and-stick*. A estrutura secundária da proteína está representada como *cartoon* verde e os resíduos chave estão representados como *sticks* com o código de três letras correspondente e o número de resíduo. Os epítipos e as regiões chave do local de ligação estão assinaladas com A, B e C. As interações chave entre os ligandos e a proteína estão representados por linhas tracejadas amarelas.

1OWE também apresenta uma conformação muito próxima da do epítipo experimental, e estabelece as mesmas interações eletrostáticas com o resíduo de ASP-192 (Figura 10c). Estes são três alvos nos quais foi apenas considerada uma região chave no local de ligação, e em que o RMSD em relação ao epítipo experimental é menor que 0.5 Å. Todos os três fragmentos estabelecem as mesmas interações intermoleculares com as regiões chave observadas no ligando experimental.

Um caso mais complexo é o do alvo 1LPZ, que possui três regiões chave de ligação, e está representado na Figura 10d-f, na qual se observam as melhores soluções que mimetizam os epítipos experimentais A, B e C deste alvo. Todos os fragmentos revelam preferência para a região definida pelo epítipo B. O fragmento equivalente ao epítipo B (Figura 10e) está numa disposição muito semelhante a este e estabelece a mesma interação eletrostática com o resíduo ASP-232. O *docking* de fragmentos para as regiões chave A e C parece ser mais difícil de prever (Figura 10d e f) e, em ambos os casos, as conformações de *docking* localizam-se na região chave B, que, como foi observado na seção anterior (Figura 7d e Figura 8e), parece “atrair” quase todos os fragmentos, tornando a amostragem das regiões chave secundárias mais difícil. Estes resultados sugerem que, em caso de alvos com mais do que uma região chave do local de ligação, a amostragem considerável de todo o local de ligação só seja possível após ocupação de uma ou mais regiões chave por um fragmento. Por outro lado alguns casos, como o do alvo 1J3J, demonstram ser possível prever a conformação de ligação de fragmentos com valores de RMSD abaixo de 1 Å em relação a dois epítipos distintos do mesmo ligando experimental. Este caso está descrito na Figura 10g e h, nas quais as conformações de *docking* dos dois fragmentos estão muito próximas de cada epítipo experimental respectivo. No alvo 1J3J, a região chave A é uma região hidrofóbica com um ambiente favorável à presença do fragmento com o anel aromático (Figura 10g) e o fragmento equivalente ao epítipo B estabelece uma interação com o resíduo de

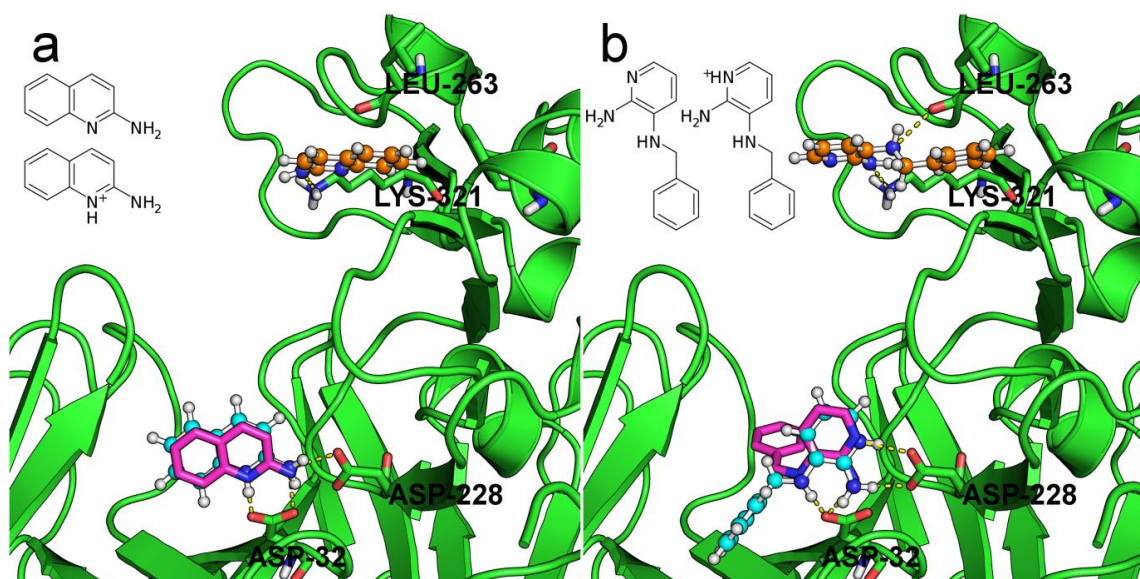


Figura 11. Conformação dos ligandos experimentais da β -secretase, com códigos PDB a) 2OHL e b) 2OHM, e a os fragmentos equivalentes da biblioteca com estados de protonação alternativos. Os ligandos experimentais estão representados como *sticks* com o código de cores seguinte: nitrogénio – azul, carbono – magenta, oxigénio – vermelho, hidrogénio – branco. Os fragmentos estão representados como *ball-and-stick*. A estrutura 2D dos diferentes estados de protonação considerados para cada fragmento está representada no canto superior esquerdo de cada painel. A estrutura secundária do alvo está representada na forma de *cartoon* verde, com os resíduos chave em *sticks* e o correspondente código de três letras e número. As interações chave entre o ligando e o alvo estão representadas como linhas tracejadas amarelas.

ASP-136 (Figura 10h). Um caso similar é o do alvo 1YWR, no qual a disposição de fragmentos equivalentes a dois, em três, epítomos é prevista com RMSD menor que 1.5 Å (Figura 9). Estes dois últimos casos, 1J3J e 1YWR, são evidências sólidas de que as bibliotecas de fragmentos e metodologias de *docking* podem ser formuladas e otimizadas de forma a prever corretamente a conformação de ligação de vários epítomos numa só experiência de *docking*. Para fechar este tópico acerca da previsão da disposição de epítomos de ligandos experimentais no local de ligação alvo, observamos que a utilização do

estado protonação correto é essencial para a previsão da conformação de ligação correta de fragmentos. Aquando da construção de uma biblioteca de fragmentos, é comum assumir que os fragmentos estão no seu estado de protonação maioritário a pH 7. No entanto, ao restringir os estados de protonação a 7, o efeito de aminoácidos acídicos ou básicos e o ambiente local que estes criam no local de ligação é ignorado. Seguindo este pressuposto, foram incluídos, durante a criação da biblioteca, os estados de protonação de todos os fragmentos entre os valores de pH de 5 e 9. A Figura 11 ilustra a conformação obtida através do *docking* de dois estados de protonação dos mesmos fragmentos da nossa biblioteca equivalentes aos fragmentos co-cristalizados com a β -secretase, com código PDB 2OHL e 2OHM. A conformação obtida com o *docking* de fragmentos dos protonados está em concordância com os complexos proteína-fragmento experimentais, estabelecendo interações iónicas com os resíduos de ASP-32 e ASP-228. Por outro lado, o *docking* do estado não protonados dos mesmo fragmentos prevê conformações totalmente distintas das experimentais.

Estes estados de protonação são justificados pela presença dos dois resíduos de aspartato (ASP-32 e ASP-228) no local ativo da enzima, e o microambiente acídico que criam. No entanto, é necessária atenção quando se consideram estes estados de protonação, dado que a sua posição no local de ligação deve justificar a sua protonação.

3.3. Classificação de fragmentos

Para além da validação das conformações de *docking* de fragmentos obtidas, o *scoring* (efetuado pelas funções de *scoring* mencionadas anteriormente) e a classificação de fragmentos segundo a sua afinidade para o alvo foram também estudados. Estes dois aspetos do *docking* requerem uma análise bastante cuidada⁶⁶. Neste estudo foram utilizadas as métricas de LE⁸¹ e FQ⁸³, e o RMSD correspondente em relação aos epítomos experimentais para classificar

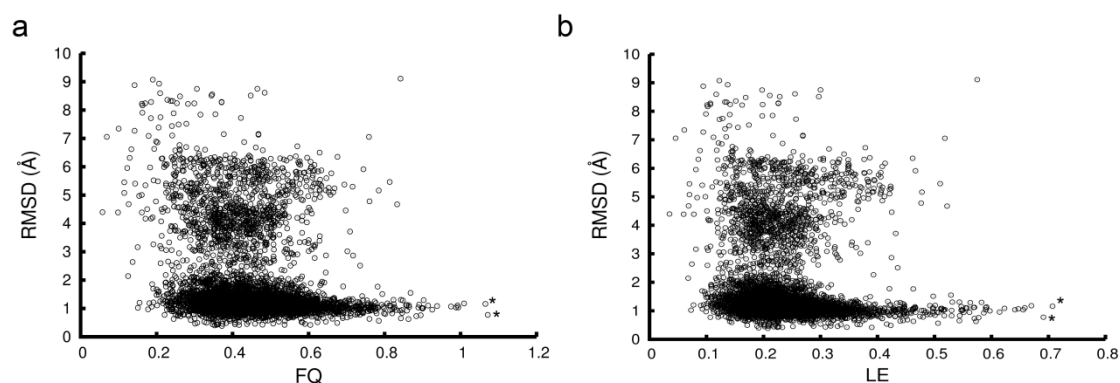


Figura 12. RMSD em relação ao epítipo em função de a) *fit quality* (FQ); b) eficiência do ligando (LE) para o alvo 1OWE. As soluções de *docking* de fragmentos com maior FQ e LE estão assinaladas com um asterisco (*) e a estrutura dos complexos está representada na Figura 13a e b.

os fragmentos. Os gráficos LE vs. RMSD e FQ vs. RMSD permitem a identificação de fragmentos com elevadas contribuições por átomo diferente de hidrogénio, específicos para cada região chave do local de ligação. A Figura 12a mostra o gráfico FQ vs. RMSD para o alvo 1OWE. É observado que os fragmentos com um valor de FQ mais elevado também possuem um valor de RMSD em relação ao epítipo do ligando experimental mais baixo, sendo que o mesmo se verifica com o gráfico LE vs. RMSD (Figura 12b). Nesta visualização podemos identificar os *clusters* de conformações principais ao longo do eixo de RMSD e, para cada *cluster*, a identificação dos fragmentos com FQ mais elevado. Os dois fragmentos com maior FQ (1.07 e 1.06) para este alvo, e que também correspondem aos fragmentos com maior LE (0.69 e 0.71 kcal/mol por átomo, respetivamente) estão representados na Figura 13a e b. Ambos os fragmentos são pequenos (com 8 e 7 átomos diferentes de hidrogénio), não aromáticos e possuem átomos de nitrogénio ionizados, os quais estão envolvidos numa interação iónica com o resíduo ASP-192, tal como acontece com o epítipo experimental. Outra interação forte que também é explorada é a ligação de hidrogénio com o resíduo de GLY-221. A classificação de fragmentos parece indicar que esta região chave tem preferência por

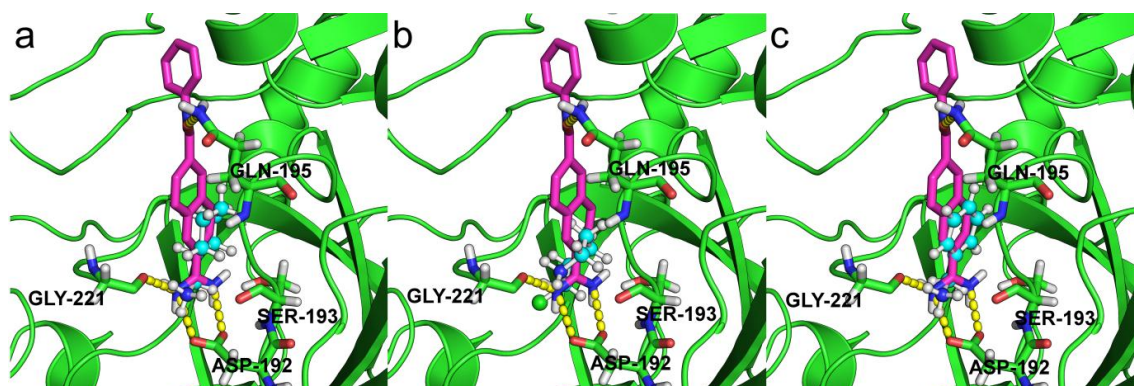


Figura 13. Conformações de ligação do ligando experimental do alvo 1OWE e de dois fragmentos com o FQ mais elevado obtido por *docking* (painéis a e b). O painel c representa a versão aromática do fragmento em a.

fragmentos não aromáticos. A versão aromática do fragmento representado na Figura 13a pode ser observada na Figura 13c, e participa nas mesmas interações que os fragmentos anteriores mas, no entanto, com um FQ e LE menores (0.78 e 0.51, respetivamente). A conformação deste último fragmento é, também, muito semelhante à do epítipo experimental. Os fragmentos do topo da classificação apresentam características estruturais similares ao epítipo do ligando experimental, o que sugere que, apesar de o *docking* não colocar os epítipo de ligandos conhecidos no topo da classificação, os fragmentos, que de facto são considerados os melhores segundo as métricas utilizadas, exploram as mesmas interações com grupos químicos similares aos observados experimentalmente.

Para além da análise FQ vs. RMSD e LE vs. RMSD, foi feita a análise de FQ e LE em função de HAC (número de átomos excluindo hidrogénios). A Figura 14a mostra a LE média em função do HAC dos fragmentos, para cada alvo. Todos os alvos revelam uma diminuição não linear de LE com o aumento do HAC, convergindo para um intervalo entre 0.18 e 0.28 kcal/mol por átomo para fragmentos de maiores dimensões. A Figura 14b mostra o caso particular do alvo 1HWI. Os valores médios de LE estão representados com um intervalo de confiança de 95%. Estes resultados mostram uma variabilidade maior

para fragmentos com menor HAC e menor variabilidade para fragmentos com maior HAC. A dependência não linear da LE com o tamanho dos fragmentos foi também descrita por Reynolds *et al*⁸³ para uma coleção de complexos proteína-ligando. O que resulta deste facto é que a classificação de fragmentos pela métrica de LE favorece fragmentos de menores dimensões, e qualquer adição de átomos durante o crescimento dos fragmentos em candidatos leva a uma diminuição brusca da LE.

A vantagem da utilização FQ em relação ao LE é permitir a comparação entre fragmentos de diferentes tamanhos, ao assumir a tendência natural da LE para diminuir com o aumento do HAC. A Figura 14c mostra um gráfico semelhante ao da Figura 14a, mas para o FQ

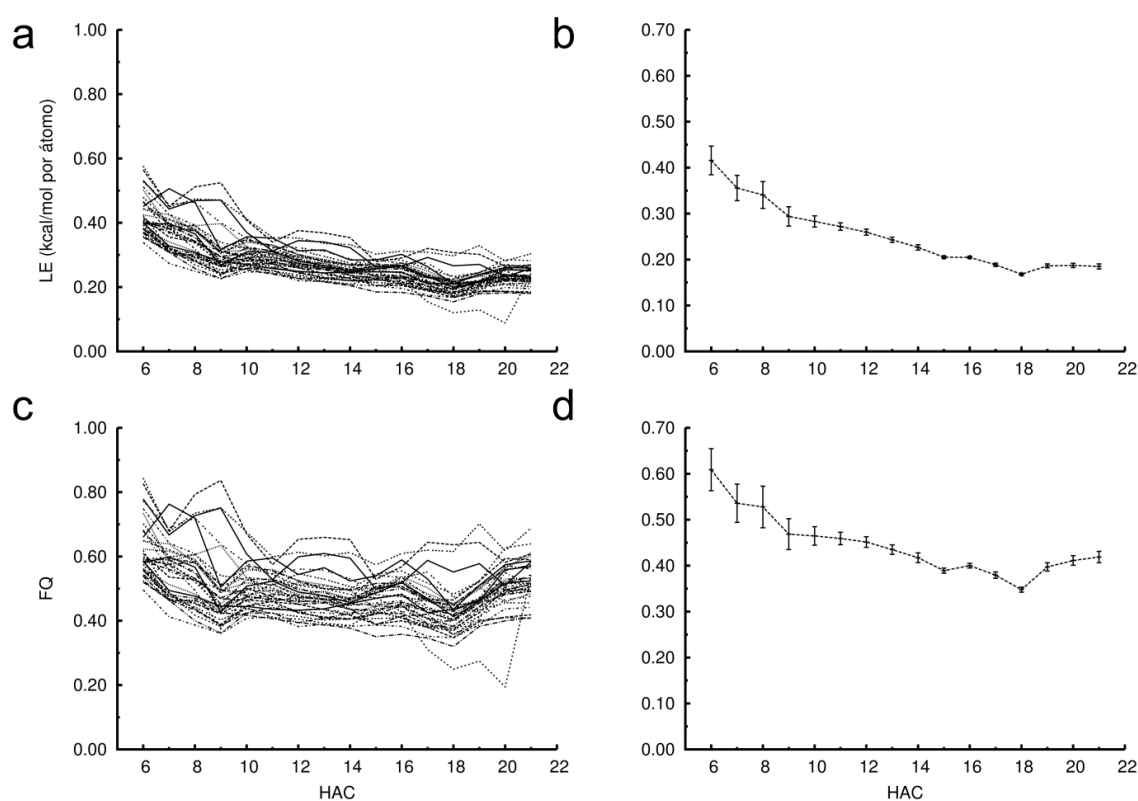


Figura 14. Média das medidas de *scoring* utilizadas em função do número de átomos diferentes de hidrogénio. Nos painéis a e b a medida é a LE para: a) todos os alvos; b) o alvo 1HWI. Nos painéis c e d a medida é o FQ para: c) todos os alvos; d) o alvo 1HWI. Nos painéis b e d são representados os intervalos de confiança de 95% para cada uma das médias apresentadas.

médio. Pode-se observar, para alguns alvos, que fragmentos mais pequenos possuem valores médios de FQ mais elevados que fragmentos maiores. Esta diferença, que também foi observada no caso do LE (Figura 14a), é mais evidente entre fragmentos com HAC entre 6 e 9. No entanto, numa visão geral, a tendência decrescente com o aumento do HAC, e que foi observada no caso do LE, não é tão evidente no caso do FQ (Figura 14c). A Figura 14d mostra a versão para FQ do gráfico da Figura 14a. Tal como no caso do LE, verifica-se uma maior variabilidade de valores de FQ para fragmentos mais pequenos, o que seria de esperar tendo em conta que o FQ é calculado a partir do LE. Neste gráfico (Figura 14d) é evidente a tendência de os fragmentos mais pequenos possuírem valores mais elevados de FQ. No entanto verifica-se que o decréscimo de FQ com o aumento de tamanho dos fragmentos é mais acentuado entre HAC de 6 e 9, sendo, após este valor, ligeiramente menos acentuado do que no caso do LE (Figura 14b). O mesmo é observável para a maioria dos alvos representados na Figura 14c, em comparação com a Figura 14a.

Estes resultados indicam que a tendência de os fragmentos mais pequenos serem favorecidos na classificação não foi eliminada pela utilização da medida de FQ na forma proposta por Reynolds⁸³. Isto pode significar que a função de *scoring* do Autodock4 sobrevaloriza as interações estabelecidas entre os fragmentos mais pequenos e alvo, o que não é surpreendente devido ao facto de esta metodologia de *docking*, tal como muitas outras, ter sido parametrizada para complexos com ligandos de tamanhos maiores do que fragmentos¹⁰⁰⁻¹⁰². Uma forma de contornar esta situação será reajustar a função de LE_{scale} , mencionada anteriormente, para resultados de *docking* de fragmentos contra uma coleção de alvos, utilizando o Autodock4.

3.4. Especificidade de fragmentos

O terceiro objetivo do trabalho é perceber o nível de promiscuidade ou especificidade dos fragmentos. Foi calculada a posição média de

cada fragmento na classificação para todos os alvos. Esta média foi então representada graficamente em função do seu desvio padrão (Figura 15). Os dados foram segmentados em quatro intervalos de HAC, 6-9, 10-13, 14-17 e 18-21 átomos, representados na Figura 15a, b, c e d, respectivamente. Estes gráficos avaliam a posição média de cada fragmento na classificação de FQ contra todos os alvos e a variabilidade da associada a esta. A Figura 15a mostra que fragmentos com HAC mais baixo (menor ou igual a 9 átomos) apresentam uma maior tendência para ocupar posições no topo da classificação do que os fragmentos com maior HAC (Figura 15d). Para além disso, os fragmentos mais pequenos e que aparecem no topo da classificação têm um desvio padrão muito baixo. A Figura 15b-c apresenta fragmentos mais distribuídos ao longo da classificação, embora os fragmentos que aparecem no topo tenham um desvio padrão baixo, tal como acontece com os fragmentos com HAC menor que 9 (Figura

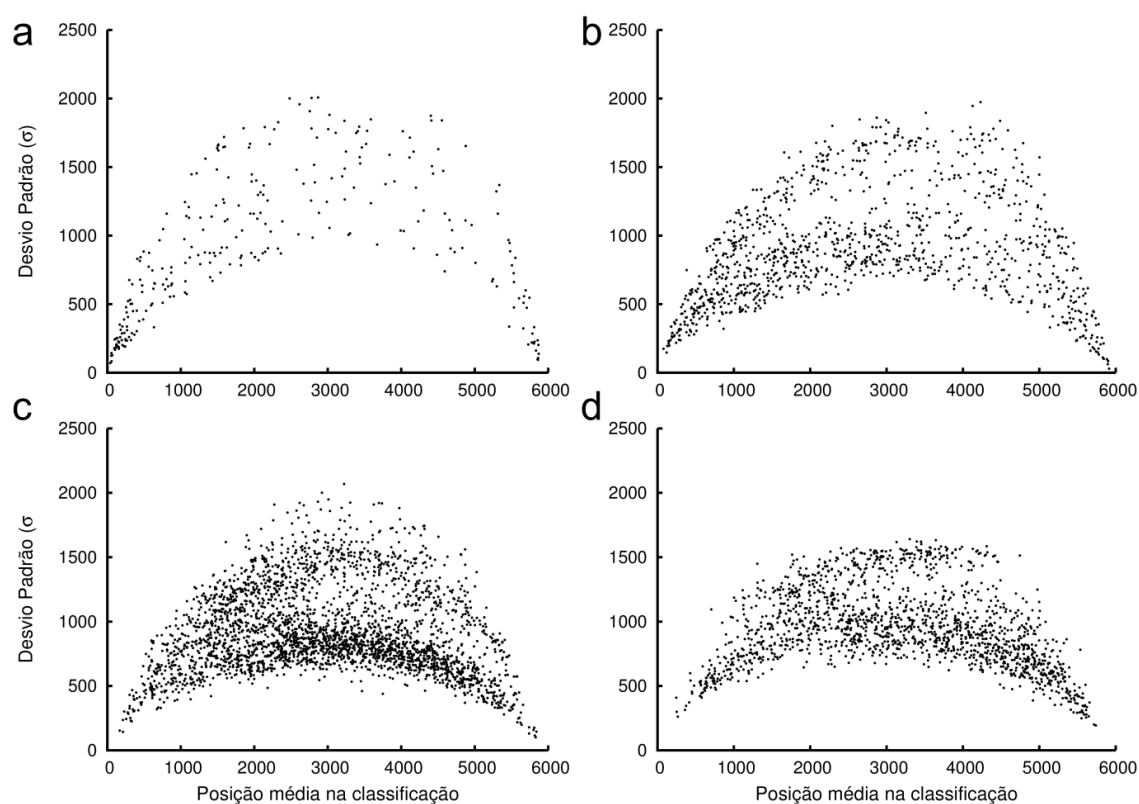


Figura 15. Posição média na classificação pela medida de FQ contra todos os alvos vs. desvio padrão da média. Os dados foram divididos em 4 intervalos de HAC: a) 6-9; b) 10-13; c) 14-17; d) 18-21.

15b). O mesmo se verifica para os fragmentos que aparecem mais bem posicionados na classificação, representados na Figura 15d. Podemos verificar assim que, em todos os intervalos de HAC considerados, existe um grupo de fragmentos – maior para os fragmentos de menores dimensões (Figura 15 a e b) – que aparece tendencialmente no topo da classificação em todos os alvos, o que é indicado pela sua posição média e pelo baixo desvio a esta associado. Estes dados sugerem que tais fragmentos são pouco específicos, ou promíscuos, para os alvos estudados.

Tal como aparecem, nos resultados de *docking*, fragmentos que estão sistematicamente no topo da classificação em todos os alvos, aparecem outros que são exatamente o oposto. Estes fragmentos aparecem sempre, em média, no fim da classificação, e com desvios padrão muito baixos (Figura 15a-d).

A utilidade, quer de fragmentos promíscuos, como de fragmentos de afinidade muito baixa (no contexto dos fragmentos) para a maioria dos alvos, numa biblioteca de fragmentos é questionável. Do ponto de vista farmacológico, uma molécula que seja promíscua terá provavelmente mais efeitos secundários, mas, no caso dos fragmentos, embora um fragmento inicial seja promíscuo, a sua evolução e otimização poderão torna-lo mais específico.

No caso dos fragmentos que ficam tendencialmente no final da classificação, estes poderão ser utilizados ser considerados controlos negativos, como forma de validação da função de *scoring*, método bastante utilizado em campanhas de *design* de fármacos com a ajuda de metodologias computacionais

Conclusões

4. Conclusões

Neste trabalho foi feita uma análise extensiva das propriedades de amostragem de uma coleção variada de alvos utilizando uma biblioteca bem definida de fragmentos moleculares, baseados em grupos químicos comuns em fármacos conhecidos. Identificamos algumas vantagens e limitações na utilização de metodologias de *docking* molecular para testar fragmentos contra alvos proteicos. Concluimos que o teste de fragmentos contra regiões chave do local de ligação alvo através de *docking* molecular pode ser utilizado para dar prioridade a fragmentos, ou classes de fragmentos, para campanhas de FBDD. Para além de dar prioridade a fragmentos, esta metodologia poderá ser utilizada para escolher as melhores regiões chave do local de ligação do alvo para “crescimento” dos fragmentos, utilizando abordagens como ligação ou evolução de fragmentos. Temos, no entanto, evidências de que as bibliotecas de fragmentos podem ser melhoradas de para a aumentar a amostragem do local de ligação de forma a obter fragmentos candidatos para todas as regiões chave do local de ligação.

A parametrização do campo de forças do Autodock4 foi capaz de colocar soluções de *docking* de fragmentos nos mesmos locais, e com orientações semelhantes aos epítipo dos ligandos experimentais conhecidos, e concluimos que pode ser utilizada em FBDD. Concluimos também que a incorporação de diferentes estados de protonação é essencial para prever corretamente a conformação de alguns fragmentos em regiões do local de ligação com características ácidas ou básicas.

Em termos da classificação dos fragmentos em termos energéticos, concluimos que a utilização da medida de *fit quality* (FQ) em vez da, mais popular, eficiência do ligando (LE), pode ser vantajosa ao permitir que fragmentos com tamanhos diferentes possam ser comparados

diretamente. É de notar que a função LE_{scale} , utilizada para o cálculo da medida de FQ, poderá ser utilizada para equilibrar funções de *scoring*, através do ajuste desta função aos resultados de docking de uma biblioteca contra uma coleção de alvos variados. A função de *scoring* do Autodock4 parece ter tendência para sobrevalorizar as energias livres de ligação de fragmentos de dimensões muito reduzidas (6 a 8 átomos diferentes de hidrogénio) e que estabeleçam interações fortes, como interações eletrostáticas envolvendo grupos carregados e ligações de hidrogénio.

Quanto à especificidade de fragmentos observámos que alguns, em quase toda a gama de tamanhos moleculares (contabilizados em termo de número de átomos diferentes de hidrogénio) da nossa biblioteca, aparecem tendencialmente no topo da classificação em todos os alvos estudados. É observável, no entanto, que destes fragmentos, ditos promíscuos, são mais abundantes os mais pequenos. Embora do ponto de vista académico, seja interessante estudar este fenómeno de “chave mestra”, do ponto de vista farmacológico, a promiscuidade é uma desvantagem pois pode significar, por exemplo, um maior número de efeitos secundários indesejados dos candidatos, e a consequente rejeição nos ensaios clínicos. Este facto levanta a questão se estes fragmentos devem ou não ser colocados numa biblioteca de fragmentos.

Num trabalho futuro será explorada a estratégia de ligação de fragmentos, verificando se será possível encontrar casos de cooperatividade positiva através de *docking* molecular.

Bibliografia

5. Bibliografia

1. Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A., Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, 70, (5).
2. Blundell, T. L., Structure-based drug design. *Nature* **1996**, 384, (6604 Suppl), 23-6.
3. Lybrand, T. P., Ligand-protein docking and rational drug design. *Curr Opin Struct Biol* **1995**, 5, (2), 224-8.
4. Verlinde, C. L.; Hol, W. G., Structure-based drug design: progress, results and challenges. *Structure* **1994**, 2, (7), 577-87.
5. Dunitz, J. D., Win Some, Lose Some - Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. *Chemistry & Biology* **1995**, 2, (11), 709-712.
6. Whitesides, G. M.; Krishnamurthy, V. M., Designing ligands to bind proteins. *Quarterly Reviews of Biophysics* **2005**, 38, (4), 385-395.
7. Williams, D. H.; Stephens, E.; Zhou, M., Ligand binding energy and catalytic efficiency from improved packing within receptors and enzymes. *Journal of Molecular Biology* **2003**, 329, (2), 389-399.
8. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J., Conformational entropy in molecular recognition by proteins. *Nature* **2007**, 448, (7151), 325-9.
9. Bissantz, C.; Kuhn, B.; Stahl, M., A Medicinal Chemist's Guide to Molecular Interactions. *Journal of Medicinal Chemistry* **2010**, 53, (14), 5061-5084.
10. Homans, S. W., Water, water everywhere--except where it matters? *Drug Discov Today* **2007**, 12, (13-14), 534-9.
11. Berne, B. J.; Weeks, J. D.; Zhou, R., Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem* **2009**, 60, 85-103.

12. Syme, N. R.; Dennis, C.; Phillips, S. E.; Homans, S. W., Origin of heat capacity changes in a "nonclassical" hydrophobic interaction. *Chembiochem* **2007**, 8, (13), 1509-11.
13. Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B., Understanding binding affinity: a combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *Journal of the American Chemical Society* **2003**, 125, (35), 10570-9.
14. Moncada, S.; Ferreira, S. H.; Vane, J. R., Prostaglandins, aspirin-like drugs and the oedema of inflammation. *Nature* **1973**, 246, (5430), 217-9.
15. Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Kjaergaard, H.; Legon, A. C.; Menucci, B.; Nesbitt, D. J., Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* **2011**, 83, (8), 1637-1641.
16. Bissantz, C.; Kuhn, B.; Stahl, M., A medicinal chemist's guide to molecular interactions. *J Med Chem* **2010**, 53, (14), 5061-84.
17. Brinck, T.; Murray, J. S.; Politzer, P., Surface Electrostatic Potentials of Halogenated Methanes as Indicators of Directional Intermolecular Interactions. *International Journal of Quantum Chemistry* **1992**, 57-64.
18. Murray, J. S.; Paulsen, K.; Politzer, P., Molecular-Surface Electrostatic Potentials in the Analysis of Non-Hydrogen-Bonding Noncovalent Interactions. *Proceedings of the Indian Academy of Sciences-Chemical Sciences* **1994**, 106, (2), 267-275.
19. Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., Halogen bonds in biological molecules. *Proc Natl Acad Sci U S A* **2004**, 101, (48), 16789-94.
20. Clark, T.; Hennemann, M.; Murray, J.; Politzer, P., Halogen bonding: the σ -hole. *Journal of Molecular Modeling* **2007**, 13, (2), 291-296.
21. Riley, K. E.; Merz, K. M., Jr., Insights into the strength and origin of halogen bonding: the halobenzene-formaldehyde dimer. *J Phys Chem A* **2007**, 111, (9), 1688-94.
22. Iltzsch, M. H.; Uber, S. S.; Tankersley, K. O.; el Kouni, M. H., Structure-activity relationship for the binding of nucleoside ligands to adenosine

- kinase from *Toxoplasma gondii*. *Biochem Pharmacol* **1995**, 49, (10), 1501-12.
23. Parks, D. J.; Lafrance, L. V.; Calvo, R. R.; Milkiewicz, K. L.; Gupta, V.; Lattanze, J.; Ramachandren, K.; Carver, T. E.; Petrella, E. C.; Cummings, M. D.; Maguire, D.; Grasberger, B. L.; Lu, T., 1,4-Benzodiazepine-2,5-diones as small molecule antagonists of the HDM2-p53 interaction: discovery and SAR. *Bioorg Med Chem Lett* **2005**, 15, (3), 765-70.
24. Benjahad, A.; Guillemont, J.; Andries, K.; Nguyen, C. H.; Grierson, D. S., 3-iodo-4-phenoxy pyridinones (IOPY's), a new family of highly potent non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg Med Chem Lett* **2003**, 13, (24), 4309-12.
25. Maillard, M. C.; Hom, R. K.; Benson, T. E.; Moon, J. B.; Mamo, S.; Bienkowski, M.; Tomasselli, A. G.; Woods, D. D.; Prince, D. B.; Paddock, D. J.; Emmons, T. L.; Tucker, J. A.; Dappen, M. S.; Brogley, L.; Thorsett, E. D.; Jewett, N.; Sinha, S.; John, V., Design, synthesis, and crystal structure of hydroxyethyl secondary amine-based peptidomimetic inhibitors of human beta-secretase. *J Med Chem* **2007**, 50, (4), 776-81.
26. Paulini, R.; Muller, K.; Diederich, F., Orthogonal multipolar interactions in structural chemistry and biology. *Angew Chem Int Ed Engl* **2005**, 44, (12), 1788-805.
27. Zhou, P.; Zou, J.; Tian, F.; Shang, Z., Fluorine bonding--how does it work in protein-ligand interactions? *J Chem Inf Model* **2009**, 49, (10), 2344-55.
28. Desiraju, G. R., Hydrogen bridges in crystal engineering: interactions without borders. *Acc Chem Res* **2002**, 35, (7), 565-73.
29. Matter, H.; Nazare, M.; Gussregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V., Evidence for C-Cl/C-Br... π interactions as an important contribution to protein-ligand binding affinity. *Angew Chem Int Ed Engl* **2009**, 48, (16), 2911-6.
30. Imai, Y. N.; Inoue, Y.; Nakanishi, I.; Kitaura, K., Cl- π interactions in protein-ligand complexes. *Protein Sci* **2008**, 17, (7), 1129-37.

31. Bohm, H. J.; Klebe, G., What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angewandte Chemie-International Edition in English* **1996**, 35, (22), 2589-2614.
32. Mecozzi, S.; Rebek, J., The 55% solution: A formula for molecular recognition in the liquid state. *Chemistry-a European Journal* **1998**, 4, (6), 1016-1022.
33. Zurcher, M.; Diederich, F., Structure-based drug design: Exploring the proper filling of apolar pockets at enzyme active sites. *Journal of Organic Chemistry* **2008**, 73, (12), 4345-4361.
34. Kuhn, B.; Hennig, M.; Mattei, P., Molecular recognition of ligands in dipeptidyl peptidase IV. *Current Topics in Medicinal Chemistry* **2007**, 7, (6), 609-619.
35. Peters, J. U.; Weber, S.; Kritter, S.; Weiss, P.; Wallier, A.; Boehringer, M.; Hennig, M.; Kuhn, B.; Loeffler, B. M., Aminomethylpyrimidines as novel DPP-IV inhibitors: A 10(5)-fold activity increase by optimization of aromatic substituents. *Bioorganic & Medicinal Chemistry Letters* **2004**, 14, (6), 1491-1493.
36. Boehringer, M.; Fischer, H.; Hennig, M.; Hunziker, D.; Huwyler, J.; Kuhn, B.; Loeffler, B. M.; Luebbers, T.; Mattei, P.; Narquizian, R.; Sebokova, E.; Sprecher, U.; Wessel, H. P., Aryl- and heteroaryl-substituted aminobenzo[a]quinolizines as dipeptidyl peptidase IV inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2010**, 20, (3), 1106-1108.
37. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A., Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, (3), 808-813.
38. Carey, C.; Cheng, Y. K.; Rossky, P. J., Hydration structure of the alpha-chymotrypsin substrate binding pocket: the impact of constrained geometry. *Chemical Physics* **2000**, 258, (2-3), 415-425.

39. Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R., C-H...pi-interactions in proteins. *Journal of Molecular Biology* **2001**, 307, (1), 357-77.
40. Burley, S. K.; Petsko, G. A., Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, 229, (4708), 23-8.
41. Burley, S. K.; Petsko, G. A., Weakly polar interactions in proteins. *Adv Protein Chem* **1988**, 39, 125-89.
42. McGaughey, G. B.; Gagne, M.; Rappe, A. K., pi-Stacking interactions. Alive and well in proteins. *J Biol Chem* **1998**, 273, (25), 15458-63.
43. Tsuzuki, S.; Uchimaru, T., Magnitude and physical origin of intermolecular interactions of aromatic molecules: Recent progress of computational studies. *Current Organic Chemistry* **2006**, 10, (7), 745-762.
44. Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K., The magnitude of the CH/pi interaction between benzene and some model hydrocarbons. *Journal of the American Chemical Society* **2000**, 122, (15), 3746-3753.
45. Celie, P. H.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K., Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **2004**, 41, (6), 907-14.
46. Rosenfield, R. E.; Parthasarathy, R.; Dunitz, J. D., Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur .1. Electrophiles and Nucleophiles. *Journal of the American Chemical Society* **1977**, 99, (14), 4860-4862.
47. Imai, Y. N.; Inoue, Y.; Yamamoto, Y., Propensities of polar and aromatic amino acids in noncanonical interactions: nonbonded contacts analysis of protein-ligand complexes in crystal structures. *Journal of Medicinal Chemistry* **2007**, 50, (6), 1189-96.
48. Palma, P. N.; Rodrigues, M. L.; Archer, M.; Bonifacio, M. J.; Loureiro, A. I.; Learmonth, D. A.; Carrondo, M. A.; Soares-da-Silva, P., Comparative

study of ortho- and meta-nitrated inhibitors of catechol-O-methyltransferase: interactions with the active site and regioselectivity of O-methylation. *Mol Pharmacol* **2006**, 70, (1), 143-53.

49. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **1997**, 23, (1-3), 3-25.

50. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* **2002**, 45, (12), 2615-2623.

51. Dudek, A. Z.; Arodz, T.; Galvez, J., Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb Chem High Throughput Screen* **2006**, 9, (3), 213-28.

52. Kubinyi, H.; Folkers, G.; Martin, Y. C., *3D QSAR in drug design: Recent advances*. Springer: 1998; Vol. 3.

53. Shirts, M. R.; Mobley, D. L.; Chodera, J. D., Alchemical free energy calculations: Ready for prime time? *Annual Reports in Computational Chemistry* **2007**, 3, 41-59.

54. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E., A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982**, 161, (2), 269-88.

55. Brooijmans, N.; Kuntz, I. D., Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **2003**, 32, 335-73.

56. Leach, A. R., *Molecular modelling: principles and applications*. Addison-Wesley Longman Ltd: 2001.

57. DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R., Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem* **1986**, 29, (11), 2149-53.

58. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996**, 261, (3), 470-89.

59. Leach, A. R.; Kuntz, I. D., Conformational analysis of flexible ligands in macromolecular receptor sites. *Journal of Computational Chemistry* **1992**, 13, (6), 730-748.
60. Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D., Flexibases: a way to enhance the use of molecular docking methods. *J Comput Aided Mol Des* **1994**, 8, (5), 565-82.
61. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **1998**, 19, (14), 1639-1662.
62. Di Nola, A.; Roccatano, D.; Berendsen, H. J., Molecular dynamics simulation of the docking of substrates to proteins. *Proteins* **1994**, 19, (3), 174-82.
63. Trosset, J. Y.; Scheraga, H. A., Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines. *Proc Natl Acad Sci U S A* **1998**, 95, (14), 8011-5.
64. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **1993**, 93, (7), 2395-2417.
65. Simonson, T.; Archontis, G.; Karplus, M., Free energy simulations come of age: Protein-ligand recognition. *Acc Chem Res* **2002**, 35, (6), 430-437.
66. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug. Discov.* **2004**, 3, (11), 935-49.
67. Bohm, H. J., LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* **1992**, 6, (6), 593-606.
68. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **1997**, 11, (5), 425-45.

69. Rognan, D.; Lauemoller, S. L.; Holm, A.; Buus, S.; Tschinke, V., Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *Journal of Medicinal Chemistry* **1999**, 42, (22), 4650-8.
70. Jencks, W. P., On the attribution and additivity of binding energies. *Proc Natl Acad Sci U S A* **1981**, 78, (7), 4046-50.
71. Arians, E. J., Speculations on the design of nonpeptidic peptidomimetics. *Trends Pharmacol. Sci.* **1982**, 3, 362-365.
72. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, 274, (5292), 1531-1534.
73. Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S.; Fong, D.; Zhu, Y. L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G., Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, (8), 3041-6.
74. Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; Burton, E. A.; Wong, B.; Tsang, G.; West, B. L.; Powell, B.; Shellooe, R.; Marimuthu, A.; Nguyen, H.; Zhang, K. Y.; Artis, D. R.; Schlessinger, J.; Su, F.; Higgins, B.; Iyer, R.; D'Andrea, K.; Koehler, A.; Stumm, M.; Lin, P. S.; Lee, R. J.; Grippo, J.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; Chapman, P. B.; Flaherty, K. T.; Xu, X.; Nathanson, K. L.; Nolop, K., Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **2010**, 467, (7315), 596-9.
75. Lipinski, C.; Hopkins, A., Navigating chemical space for biology and medicine. *Nature* **2004**, 432, (7019), 855-61.

76. Walters, W. P.; Stahl, M. T.; Murcko, M. A., Virtual screening--an overview. *Drug Discovery Today* **1998**, 3, (4), 160-178.
77. Chen, Y.; Shoichet, B. K., Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, 5, (5), 358-64.
78. Hann, M. M.; Leach, A. R.; Harper, G., Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* **2001**, 41, (3), 856-64.
79. Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R., Fragment-based lead discovery. *Nature Reviews Drug Discovery* **2004**, 3, (8), 660-72.
80. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **2003**, 8, (19), 876-7.
81. Hopkins, A. L.; Groom, C. R.; Alex, A., Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* **2004**, 9, (10), 430-1.
82. Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C., Fragment-based lead discovery: leads by design. *Drug Discov Today* **2005**, 10, (14), 987-92.
83. Reynolds, C. H.; Bembenek, S. D.; Tounge, B. A., The role of molecular size in ligand efficiency. *Bioorganic & Medicinal Chemistry Letters* **2007**, 17, (15), 4258-61.
84. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K., BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **2007**, 35, (Database issue), D198-201.
85. Alex, A. A.; Flocco, M. M., Fragment-based drug discovery: what has it achieved so far? *Current Topics in Medicinal Chemistry* **2007**, 7, (16), 1544-67.
86. Congreve, M.; Murray, C. W.; Blundell, T. L., Structural biology and drug discovery. *Drug Discov Today* **2005**, 10, (13), 895-907.
87. Mannhold, R.; Kubinyi, H.; Folkers, G.; Jahnke, W.; Erlanson, D. A., *Fragment-based approaches in drug discovery*. Wiley-VCH: 2007.

88. Rohrig, C. H.; Loch, C.; Guan, J. Y.; Siegal, G.; Overhand, M., Fragment-based synthesis and SAR of modified FKBP ligands: influence of different linking on binding affinity. *Chemmedchem* **2007**, 2, (7), 1054-70.
89. Bohm, H. J.; Banner, D. W.; Weber, L., Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J Comput Aided Mol Des* **1999**, 13, (1), 51-6.
90. Liebeschuetz, J. W.; Jones, S. D.; Morgan, P. J.; Murray, C. W.; Rimmer, A. D.; Roscoe, J. M.; Waszkowycz, B.; Welsh, P. M.; Wylie, W. A.; Young, S. C.; Martin, H.; Mahler, J.; Brady, L.; Wilkinson, K., PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *Journal of Medicinal Chemistry* **2002**, 45, (6), 1221-32.
91. Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H., Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. *Journal of Medicinal Chemistry* **2001**, 44, (26), 4615-27.
92. Verdonk, M. L.; Giangreco, I.; Hall, R. J.; Korb, O.; Mortenson, P. N.; Murray, C. W., Docking performance of fragments and druglike compounds. *J. Med. Chem.* **2011**, 54, (15), 5422-31.
93. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H., Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, 48, (2), 403-13.
94. Bemis, G. W.; Murcko, M. A., The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, 39, (15), 2887-93.
95. Bemis, G. W.; Murcko, M. A., Properties of known drugs. 2. Side chains. *J. Med. Chem.* **1999**, 42, (25), 5095-9.
96. ChemAxon – cheminformatics platforms and desktop applications. www.chemaxon.com

97. Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General Atomic and Molecular Electronic-Structure System. *J. Comput. Chem.* **1993**, 14, (11), 1347-1363.
98. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993**, 97, (40), 10269-10280.
99. Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W., Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, 50, (4), 726-741.
100. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, 19, (14), 1639-1662.
101. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S., A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* **2007**, 28, (6), 1145-52.
102. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* **1996**, 10, (4), 293-304.
103. Gasteiger, J.; Marsili, M., Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, 36, (22), 3219-3228.
104. Marsili, M.; Gasteiger, J., Pi-Charge Distribution from Molecular Topology and Pi-Orbital Electronegativity. *Croat. Chem. Acta* **1980**, 53, (4), 601-614.
105. W.L.DeLano, *The PyMOL Molecular Graphics System*. in: D. Scientific (Ed.), Palo Alto: USA, 2002.

Anexos

6. Anexos

Anexo I (página seguinte)- Número de soluções de *docking* de fragmentos em diferentes intervalos de RMSD em relação a um determinado epítopo experimental para todos os alvos estudados. O código de referência PDB identifica cada gráfico. Os dados de RMSD para alvos com mais do que um epítopo experimental estão representados com o seguinte esquema de cores: epítopo A – vermelho; epítopo B – verde; epítopo C - azul.

